UNIVERSITY OF SANTIAGO DE COMPOSTELA
DEPARTMENT OF FINANCIAL ECONOMICS AND ACCOUNTING

UNIVERSIDAD DE SANTIAGO DE COMPOSTELA
DEPARTAMENTO DE ECONOMÍA FINANCIERA Y CONTABILIDAD

Are the ratings useful tools selecting mutual funds?

¿Los ratings son herramientas útiles en la selección de fondos de inversión?

Renato Heitor Correia Domingues

Santiago de Compostela 2017

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

UNIVERSIDAD DE SANTIAGO DE COMPOSTELA

DEPARTAMENTO DE ECONOMÍA FINANCIERA Y CONTABILIDAD

Are the ratings useful tools selecting mutual funds?

¿Los ratings son herramientas útiles en la selección de fondos de inversión?

Tesis para la obtención del grado de doctor, presentada por el licenciado Don Renato Heitor Correia Domingues, la cual fue realizada en el Departamento de Economía Financiera y Contabilidad de la Universidad de Santiago de Compostela, bajo la dirección de Don Luis Alberto Otero González, profesor titular de Economía Financiera y Contabilidad y Don Pablo Duran Santomil, profesor ayudante doctor de Organización de Empresas y Comercialización de la Universidad de Santiago de Compostela.

Vº Bº

D. Luis Alberto Otero González

D. Pablo Duran Santomil

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

D. LUIS ALBERTO OTERO GONZÁLEZ, profesor titular del Departamento de Economía Financiera y Contabilidad de la Universidad de Santiago de Compostela y D. PABLO DURAN SANTOMIL, profesor ayudante doctor de Organización de Empresas y Comercialización de la Universidad de Santiago de Compostela

CERTIFICAMOS que D. Renato Heitor Correia Domingues ha realizado bajo nuestra dirección el trabajo de investigación "¿Los ratings son herramientas útiles en la selección de fondos de inversión?". Este trabajo reúne las condiciones necesarias para ser presentado y juzgado como tesis doctoral y, por tanto, para optar al grado de doctor en Ciencias Económicas y Empresariales.

Santiago de Compostela, 2017

Fdo. Prof. Dr. Luís A. Otero González
Profesor titular del Departamento de
Economía Financiera y Contabilidad
Universidad de Santiago de Compostela

Fdo. Prof. Dr. Pablo Durán Santomil
Profesor Ayudante Doctor del
Departamento de Organización de
Empresas y Comercialización
Universidad de Santiago de Compostela

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

For my family, especially for Renata

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

Acknowledgement

My journey through this thesis was long and it would not be possible without the support of several people. First and foremost, I would like to express my gratitude to Luis Alberto Otero González for his fantastic supervision. His encouragement, patience and guidance were vital for my development as a researcher. To Pablo Duran Santomil, for his co-supervision that helped me to improve time management, technical details and quality standards. It is hard to find the words to express how grateful I feel for their work in this project.

To Nuno Miguel Cruz Domingues for encouraging me to go forward with this PhD, for changing his critical opinion and for participating in this study's debate.To all my dear family for their outstanding support and endless encouragement that helped to nurture a positive attitude throughout the challenges. To Marta Venade, for her friendship and help with the English revision.

Last but not least, a big thank you to all my PhD Professors and their fellow students for their availability. Thank you very much again, it has been a pleasure to work with you all.

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

Abstract

This thesis explores three topics that contribute to understanding how ratings can help investors selecting mutual funds. Specifically, this thesis pretends understand the ability to detect funds which outperform their peers based on Morningstar Star Rating, Morningstar Analyst Rating and Morningstar Sustainability Scores. The chapters are independent of each other in terms of theoretical grounding and methodology but complement each other by investigating the three different angles related.

In Chapter 2 is analysed if the Morningstar Star Rating can have the predictive power of future performance, as well how another's variables mixed with these ratings can help investors making better choices for increase their wealth and minimize their risk. In the Chapter 3 the roll of qualitative ratings is studied, more specifically the Morningstar Analyst Rating. In particular, we analyse if the ratings based on analyst opinions can help identifying mutual funds that outperform their peers in the short and long time in terms of risk adjusted performance, and if the addition of other variables can improve the selecction. Of particular interest is the combination of quantitative ratings (Star Rating) and qualitative (Analyst) to improve the selection. In Chapter 4 Morningstar Sustainability Scores and ESG (environmental, social and governance) scores are used to evaluate the effect of portfolio management decisions based on sustainable criteria.

The objective of the Chapter 2 is to analyse if ratings are reliable tools in selecting mutual funds. Our sample contains all European equity funds rated by Morningstar from 2004 to 2014. Our conclusions support the ability of ratings to identify outperformers in the short and medium term. In this sense, we have found that on average, funds lower rated have a worst out of sample performance in terms of risk adjusted measures and annual returns. The strongest predictability is observed for one year ahead, but is also good for the three years' period. The inclusion of costs and other variables like turnover, age and size, reflect the importance to consider other factors to explain future performance. Finally, the best ratings have a better behaviour in terms of VaR (value at risk) for the 99% confidence levels, showing that the investment in good rated funds can help to preserve better the investors´ wealth under unfavourable market conditions.

The objective of Chapter 3 is to evaluate the usefulness of analyst assesments to select investments, using the Morningstar Analyst. Morningstar Analyst Ratings are forward-looking qualitative and quantitative analyses of mutual fund about five pillars: Process, Performance, People, Parent and Price, that includes factors like the cost, past performance, quality of

management, interest alignment, etc. In this chapter, we assess to what extent selecting mutual funds based on Morningstar Analyst and Star ratings criteria has an impact on the performance of investors. In particular, we try to answer if good analyst ratings outperform non-recommended ones in the short (12 month) and long term (36 month) and if it is useful to combine both ratings in the screening process to identify good future performers. Our sample contains US equity funds covering the period August 2012 to August 2016. Our conclusions support the ability of Gold ratings to select funds that will behave better in terms of future performance. We have found little evidence that, on average, funds with a better Analyst Rating (Gold) have a better performance in terms of risk adjusted measures (alpha and sharpe). The predictability is observed in several analyses done in one year ahead but not for three-years. This evidence is more relevant in the case of the analysis made by investment style´s category. The combination of Stars and Analyst ratings does have medium-term differentiation results, with a higher performance in terms of Sharpe's ratio for bronze (3, 4 or 5 stars) or three-star gold funds.

The objective of Chapter 4 is to evaluate the effect of sustainability scores and the different dimensions in which ESG score is subdivided (Environmental, Social and Governance) on performance and cash flow. On the other hand, the conventional dichotomous variable has been added to the models to evaluate to what extent the results may be different. Our result shows that there are a large number of funds that are not declared sustainable but their portfolio is comparable to sustainable mutual funds. Furthermore, Morningstar Sustainability Score is significant explaining the level of performance for all the metrics analysed (alpha, Sharpe and annual return), with negative sign in most models. In terms of downside risk, the level of sustainability is negatively and significantly related with the VaR of the fund, supporting that higher scored mutual funds protect better against extreme losses.

Resumen

Esta tesis analiza tres temas que son de interés general para inversores, analistas e investigadores: ratings cuantitativos, ratings cualitativos y puntuaciones de sostenibilidad. La primera cuestión versa sobre la importancia de los ratings cuantitativos de los fondos de inversión de cara a la toma de decisiones de inversión. Los ratings cuantitativos establecen una clasificación de los fondos en base a su comportamiento pasado de rentabilidad- riesgo, de forma que los mejores fondos son mejor valorados. Dado que muchos inversores emplean en la actualidad estos ratings a la hora de la toma de decisiones de inversión, el primer capítulo se centra en analizar si dichos fondos obtienen mejores resultados futuros o un menor riesgo. La segunda cuestión de investigación que se analiza es si los ratings cualitativos, es decir aquellos que analizan aspectos como la firma gestora o el equipo gestor, tienen poder predictivo sobre la performance o el riesgo de los fondos de inversión. Finalmente, dada la importancia que están adquiriendo las inversiones socialmente responsables (SRI), se analiza el efecto que sobre rendimientos, flujos y riesgo tienen los fondos con mejores puntuaciones de sostenibilidad. La tesis aborda cada uno de estos tres temas en tres capítulos distintos, en los que en cada uno de ellos se realiza un análisis empírico para aportar conclusiones, empleando en todos ellos la misma base de datos Morningstar Direct.

En la literatura especializada está documentada la influencia de los cambios en los ratings cuantitativos en la entrada y salida de los flujos de los fondos de inversión por parte de los inversores. Esto indica que el inversor está preocupado de los ratings de los fondos y de sus variaciones.Sin embargo, lo que no está tan claro en la literatura, ya que hay autores que encuentran evidencia en contra y otros a favor, es si dichos ratings tienen el poder de detectar la habilidad de los gestores de los fondos y por tanto de predecir rentabilidades futuras. Los principales ratings cuantitativos que emplean los inversores para la toma de sus decisiones son los ratings de estrellas de Morningstar. Su metodología consiste en asignar de 1 (peor) a 5 (mejor) estrellas cada fondo con base en su posición respecto a otros fondos comparables o de su misma categoría en relación a su rentabilidad ajustada al riesgo histórica. Así, en el Capítulo 2 se analiza el poder predictivo de los ratings cuantitativos de Morningstar, es decir, si contribuyen a entender la habilidad de los gestores de fondos de inversión en batir al mercado, tema conocido en la literatura como persistencia. De esta forma, este capítulo trata de demostrar si los fondos con mejores ratings cuantitativos de Morningstar obtienen una mejor *performance* futura en términos de diversos indicadores de rendimientos ajustados y no ajustados al riesgo. Otra cuestión que se plantea en el análisis de este capítulo es si los inversores que invierten basándose

en los ratings cuantitativos de Morningstar, es decir, eligiendo los fondos con mejor rating, estarán más protegidos en condiciones de mercado adversas que los que eligen fondos con peor rating. Para ello se analizará el efecto que tienen los ratings sobre el VaR o valor en riesgo de la inversión.

En el primer capítulo usamos una muestra de 1,579 fondos europeos de renta variable, de 2004 a 2014, que disponen de rating de Morningstar global (*Overall*), que es una media ponderada de los ratings a 3, 5 y 10 años. Los fondos de inversión seleccionados son los denominados fondos abiertos, seleccionando una única clase para cada uno de los fondos de inversión disponibles. Para llevar a cabo nuestra investigación empleamos dos medidas de rentabilidad ajustadas al riesgo (Alpha de Jensen y ratio de Sharpe) y una sin ajustar (rendimientos anuales netos del fondo). Para analizar el riesgo en lo relativo a pérdidas máximas, empleamos la métrica de valor en riesgo o VaR con un 99 % de probabilidad.

La primera metodología adoptada para verificar si los fondos con mejores Morningstar Star Ratings obienen mejor rentabilidad futura ha sido mediante técnicas de datos de panel. Esta metodología se emplea con retardos de 1 y 3 años para evaluar el poder predictivo de dichos ratings cuantitativos en distintos plazos. Al análisis se han añadido como variables de control los gastos, el tamaño o la experiencia del gestor. Posteriormente también empleamos regresión cuantílica, basada en Chen and Huang (2011), para analizar el posible efecto de distintos cuantiles sobre las diferentes métricas de desempeño.

Las conclusiones de este primer estudio evidencian una relación positiva entre la calificación del fondo y la rentabilidad futura. Los fondos de inversión con mejores calificaciones tienen, en media, mejor rendimiento que los que tienen peores calificaciones, tanto para las medidas ajustadas al riesgo, como también, para los rendimientos netos anuales. Nuestros resultados están en línea con trabajos como los de Morey y Gottesman (2006), Müller y Weber (2014), y Meinhardt (2014). Este poder predictivo lo observamos a un año vista, pero también es válido para un período de tres años. La inclusión de otras variables como costes, tamaño o edad refleja la importancia de considerar otras variables en la selección de fondos de inversión. Sin embargo, las calificaciones siguen siendo significativas para explicar el desempeño. Finalmente, las mejores calificaciones tienen un mejor desempeño en términos de valor en riesgo (VaR), mostrando que la inversión en fondos de buena calificación puede ayudar a preservar mejor la riqueza de los inversores ante caídas del mercado.

Las calificaciones más populares para los inversores de fondos son las calificaciones cuantitativas, en especial los ratings de estrellas de Morningstar. No obstante, la selección de

fondos basada exclusivamente en la performance histórica o ratings cuantitativos excluye todo un conjunto de factores cualitativos que pueden ser mejores indicadores de la performance futura o ayudar con los aspectos cuantitativos a predecir mejor la rentabilidad futura. En el capítulo 3 analizamos los ratings cualitativos de Morningstar. Los ratings cualitativos son relativamente recientes, por lo que existen actualmente un número reducido de estudios que los analizan y sobre muestras temporales cortas. Morningstar cuenta actualmente con dos tipos de ratings cualitativos: los denominados Stewardship y los Analyst. Nuestro estudio se centra en los Morningstar Analyst Ratings, ya que los Stewardship Grade no están disponibles más que para fondos de inversión de Norteamérica por lo que ni es tan extendido su uso ni dispondríamos de una muestra lo suficientemente elevada para obtener conclusiones robustas. Los Analyst Ratings son una calificación reciente creada por Morningstar en 2011 que se basa en análisis de aspectos cualitativos de los fondos de inversión. Estos ratings se expresan como medallas: Oro, Plata, Bronce, Neutral y Negativo. Las medallas Oro, Plata y Bronce son las notas que se atribuyen a los mejores fondos, mientras que Neutral y Negativo son notas a los peores fondos. Estas calificaciones difieren mucho de las calificaciones de las estrellas, que son medidas basados en la rentabilidad pasada de un fondo en su categoría. Aunque dicho rating no supone una recomendación de compra, se ha creado con una convicción por parte de Morningstar de que tiene poder prospectivo, de esta forma un fondo calificado como oro tiene el más alto nivel de convicción, mientras que un fondo con calificación negativa probablemente tendrá un desempeño significativamente inferior a otros fondos comparables (de la misma categoría). La calificación está basada en la nota que obtiene en los denominados cinco pilares: personas, precio, proceso, gestora y performance. Existen muy pocos estudios que se examinan los ratings cualitativos y su poder para predecir la *performance* futura (Wellman y Zhou, 2007; Seng, 2009; Gottesmann y Morey, 2012; Kamal, 2013; Haslem, 2014; Armstromg, Genc y Verbeek, 2016). En general, sus resultados han demostrado la incapacidad de predecir el rendimiento futuro.

Para llevar a cabo este trabajo empleamos la base de datos de Morningstar Direct. Esta base contiene datos desde 2011 para fondos con calificación de Analyst, pero la muestra es muy limitada hasta la mitad de 2012, cuando hay alrededor de 200 fondos de inversión calificados para el área de inversión de EEUU. Algunos trabajos anteriores han incluido todo el universo de fondos de inversión con calificación, a pesar de la heterogeneidad en términos de área de inversión, riesgo de tipo de cambio, período de análisis, entre otros. Por eso en nuestra investigación sólo incluimos las calificaciones de Analyst para fondos a partir de agosto 2012 a agosto de 2016. Limitamos nuestra selección a los fondos centrados en el área de inversión de Estados Unidos e incluidos en las nueve categorías comunes resultantes de la combinación de

tamaño y valor. La muestra contiene 10.772 observaciones mensuales y un promedio de 220 fondos de inversión clasificados, con una buena representación en cada nivel excepto en el caso negativo. Siguiendo a Amstrong *et al.* (2016) distinguimos entre fondos "no recomendados" y "recomendados", donde los "recomendados" están compuestos por las categorías de oro a bronce, y neutral y negativo se clasifican en "no recomendado". En general, más del 80% de los fondos son clasificados como "recomendados", y esto es explicado por Morningstar por el hecho de que priorizan fondos de alta calidad y en general son más grandes y con menores costes y tasas de rotación.

Para el análisis empírico, hemos estimado distintos modelos para evaluar el rendimiento en base los ratings Analyst y sus pilares. Dado que es un rating con carácter prospectivo (*fordward looking*) nuestra hipótesis de investigación es que reflejara el rendimiento futuro en el largo plazo, de forma los fondos recomendados obtendrían mayor rendimiento futuro. Se calcula que no se muestra el tiempo de retorno de la respuesta por 12 y 36 meses después de la inicialización de la calificación y, a continuación, devolver las diferentes métricas de performance utilizando oro, plata y bronce añadiendo después variables de control. Utilizamos lo método basado en los datos de panel que puede controlar los efectos individuales de los efectos de la colinealidad y la eficiencia, entre otros (Baltagi, 2010).

A semejanza del capítulo anterior, los modelos de panel fueron inicialmente estimados sin variables de control para evaluar el efecto de seleccionar fondos basados exclusivamente en el rating Analyst. Posteriormente, se han incluido otras variables explicativas tales como costes, tamaño y años. En todos los casos se han incluido variables de control la categoría del fondo y el efecto temporal de la rentabilidad. Si bien los resultados dependen de la métrica utilizada, en general, sólo los fondos clasificados como Oro muestran un mejor desempeño en un horizonte de 12 meses sobre los fondos "no recomendados" en términos de la ratio de Sharpe. A 36 meses los resultados son más decepcionantes. Por lo tanto, nuestro análisis revela la incapacidad de las calificaciones de los analistas para identificar los fondos que superan a sus pares excepto en el caso de los fondos de oro, donde los resultados muestran un mejor desempeño que los clasificados como no recomendados a corto plazo. Las diferencias con estudios previos pueden deberse al hecho de considerar en el estudio una muestra diferente centrada exclusivamente en los EE.UU., la metodología de datos de panel y el período considerado.

Posteriormente hemos incluido los pilares en los que los ratings Analyst se desglosan para analizar si el rendimiento futuro está relacionado con cualquier dimensión específica. El hecho de que haya muy pocos fondos de inversión con un pilar negativo conduce a una

comparación entre pilares fundamentalmente positivos y neutrales. Los resultados obtenidos muestran, en general, coeficientes no significativos, con signo negativo y positivo. La alta subjetividad que puede conducir al proceso de evaluación y la dificultad para establecer un umbral entre una evaluación positiva y otra neutral pueden explicar los resultados obtenidos.

Para entender si la combinación de ambas calificaciones (cuantitativas y cualitativas) pueden ayudar en el proceso de identificación de los mejores resultados, también estudiamos el efecto de considerar conjuntamente los dos tipos de ratings. Así, se crearon diferentes variables que resultan de la combinación del mejor analista (Oro a Bronce) y estrellas (4 o 5 estrellas) y considerando el resto como "no recomendado. Los resultados del análisis conjunto de los fondos muestran que cuando las calificaciones se utilizan aisladamente, sólo los fondos de cuatro o cinco estrellas y los fondos Oro superan a la clasificación cuantitativa más baja a corto plazo. Cuando ambos criterios se combinan los fondos Oro y cuatro o cinco estrellas tienen un mejor desempeño a 12 meses, pero no a 36 meses. Esto significa que los inversores que basan sus decisiones en ambos criterios deben monitorear las carteras anualmente y verificar que continúan mantienen ambas calificaciones. Por otro lado, la combinación de las dos calificaciones tiene resultados a medio plazo, con un mayor rendimiento en términos de Sharpe para los fondos Bronce de 3, 4 o 5 estrella y Oro de tres estrellas, presentando sólo un signo negativo los fondos Plata de tres estrellas.

Para dar más consistencia al estudio, hicimos tests complementarios de robustez. De esta forma utilizamos la regresión cuantil para extender el modelo de regresión a los cuantiles condicionales de las diferentes métricas de rendimiento, ya que está técnica puede ser más apropiada para un universo heterogéneo de fondos de inversión donde las estrategias y los objetivos pueden variar (Chen y Huang, 2011). Este modelo nos permite capturar información sobre los coeficientes en diferentes cuantiles de la variable dependiente dado el conjunto de variables endógenas. Los resultados de la regresión cuantil muestran que, en general, los signos no son significativos para la mayoría de las calificaciones Analyst, siendo los fondos Oro los únicos que superan a los “no recomendados” en algunos cuantiles. Cuando se incluye el efecto de los costes o el tamaño de los fondos, las calificaciones Oro que son significativas dejan de serlo. Por lo tanto, es razonable pensar que tanto el tamaño de los fondos como los costes pueden explicar las diferencias de desempeño más que las calificaciones de los analistas.

En el Capítulo 4, nuestro estudio se centra en el nuevo rating de sostenibilidad de Morningstar (*Morningstar Sustainability Rating*), en concreto emplearemos los puntuaciones o *scores* de Sostenibilidad y de ESG (iniciales de *Environmental*, *Social* y *Governance*). El

objetivo de este capítulo es extender el tradicional debate de si las inversiones en fondos de inversión sostenibles afectan a la *performance* de un forma positiva, neutra o negativa. Las inversiones socialmente responsables (SRI) en fondos de inversión es una opción que tiene en cuenta los criterios ambientales, sociales y de gobierno corporativo (ESG) para generar valor a largo plazo. Otros nombres asociados a este tipo de inversión en la literatura son el de inversión social, sostenible, socialmente consciente, verde, responsable o ética. Los tradicionales estudios realizados hasta la fecha han empleado una variable dicotómica para valorar si un fondo es socialmente responsable o no con base en cómo la gestora lo declara en su Folleto de Gestión. Sin embargo, la reciente publicación en el año 2016 de Morningstar de los *ratings* y puntuaciones de sostenibilidad nos permite analizar las diferencias, en vez de aplicando un criterio binario (si/no) las puntuaciones alcanzadas por cada fondo.

Varios estudios han concluido que las empresas que tienen políticas y prácticas de responsabilidad social son buenas inversiones y, por lo tanto, los fondos que invierten en ellas. Recientemente, por ejemplo, Friede, Busch y Bassen (2015) llevó a cabo un meta análisis de unos 2.200 estudios empíricos, convirtiéndolo en la revisión más completa de la investigación académica sobre este tema. Encontraron que la mayoría de los estudios muestran una correlación positiva entre los factores ESG y el rendimiento financiero. Pero incluso a pesar de las investigaciones llevadas a cabo hasta la fecha todavía hay un debate acerca de si este tipo de inversiones pueden crear valor para los inversores o no. Aunque según Lewis y Mackenzie (2000) y Webley, Lewis y Mackenzie (2001) algunos inversores en fondos SRI están dispuestos a aceptar rendimientos menores por su postura moral, el desempeño de los fondos SRI y convencionales es una cuestión todavía abierta, especialmente cuando se analizan las puntuaciones de sostenibilidad de un fondo, ya que hasta la fecha sólo ha sido analizado por El Ghoul y Karoui (2017).

En la literatura que realiza estudios con variables dicotómicas para diferenciar los fondos SRI de los fondos convencionales se han encontrado evidencias contrapuestas de si éstos obtienen mejores resultados. Autores como Junkus y Berry (2015) sustentan que las inversiones en fondos SRI tienen desempeño muy similar que los fondos convencionales. Estudios como Luther, Mattako y Corner (1992) y Mallin, Saadouni y Briston (1995) apoyan la idea de que los fondos SRI superan a los índices del mercado. Pero la teoría más convencional es que los fondos mutuos de ISR tienen el mismo rendimiento que los fondos de otros, y autores como Hamilton, Jo y Statman (1993), Kreander, Gray, Power y Sinclair (2002, 2005), Gregory y Whitakker Bauer, Derwall y Otten (2007) o Humphrey, Warren y Boon (2016) están en línea con esta teoría. Nuestro estudio aporta como principal novedad a la actual literatura sobre la rentabilidad

de los fondos de inversión socialmente responsables el examen del efecto del grado de sostenibilidad, medido a través de las puntuaciones de sostenibilidad y de ESG de Morningstar.

Las puntuaciones de sostenibilidad y ESG de Morningstar se elaboran para cada fondo que invierte al menos el 50% de sus activos en compañías con puntuaciones calculadas por Sustainalytics. Sustainalytics es una compañía líder en las valoraciones de criterios ESG de compañías de todo el mundo. La metodología de ESG de Sustainalytics consiste en distintos indicadores ESG, de 120 a 150 en función de la industria, para medir las prácticas sostenibles de una empresa. Sustainalytics evalúa el desempeño de la empresa en cada uno de estos indicadores a partir de varias fuentes de datos internas y externas (Sustainalytics, 2016). Estos indicadores se engloban en los tres pilares de la puntuación ESG, que son medio ambiente, social y gobernanza. En cada pilar se distinguen varias categorías de indicadores: básicos y específicos del sector. De esta forma para cada compañía obtiene mediante agregación de los indicadores una puntuación para cada uno de los pilares ESG, y por medio de las denominadas controversias, de la puntuación global de sostenibilidad. Morningstar agrega las puntuaciones de los activos que forman la cartera de un fondo de inversión de forma normalizada para obtener la puntuación total del fondo en cada pilar ESG, la puntuación total de ESG, las controversias del fondo, y finalmente, por diferencia entre la puntuación ESG y las controversias, la puntuación de sostenibilidad de un fondo de inversión, de forma que este puede obtener un valor máximo teórico de 100.

Nuestro estudio tiene una muestra inicial de 1.593 fondos de inversión de renta variable europea de tipo abierto (*open funds*) con puntuaciones de sostenibilidad de Morningstar en noviembre de 2016. La selección de fondos se ha hecho evitando problemas de multicolinealidad, por lo que se ha selecciona sólo una clase. Para cada fondo se analiza el efecto que tienen las puntuaciones de sostenibilidad, ESG y de cada uno de los pilares de forma independiente en diferentes medidas de desempeño. Además, se ha tenido en cuenta otras variables como el tamaño, los gastos y la antigüedad del fondo. El número final de fondos varió cuando se consideran los costes, de forma que la muestra se reduce de 1.593 a 571 motivados por la falta de datos en Morningstar Direct para todas las variables consideradas. Hemos analizado el rendimiento y el efecto de riesgo utilizando las métricas de rentabilidad y riesgo de los últimos dos años basados en el trabajo de Wimmer (2012), que muestra que las puntuaciones ESG persisten durante dos años motivadas por los cambios en las tenencias de activos de los fondos de inversión SRI. En particular, como medidas de performance empleamos la rentabilidad anual, el alfa de Jensen y ratio de Sharpe. Además, se ha considerado el efecto que la sostenibilidad tiene en los flujos de efectivo de un fondo y también su efecto sobre el valor en riesgo (VaR).

Los resultados obtenidos muestran que hay un gran número de fondos que no se declaran sostenibles, pero su cartera es comparable a los fondos de inversión que se declaran sostenibles. Además, el parámetro asociado en la regresión a la puntuación de sostenibilidad es significativo, explicando el nivel de desempeño de todas las métricas analizadas (alfa, Sharpe y rendimiento neto anual), con signo negativo en la mayoría de los modelos. Utilizando una variable *dummy* convencional para identificar los fondos de inversión SRI, los resultados son significativos, pero, al contrario, mostrando que considerar el nivel de sostenibilidad puede ayudar a entender mejor la relación entre desempeño y responsabilidad social. Nuestros resultados están de acuerdo con Statman y Glushkov (2016), quienes concluyen que la falta de criterios claramente definidos para distinguir los fondos SRI afecta a los resultados de los estudios previos de la literatura, y eso puede explicar el por qué se encuentran en la misma resultados contradictorios. Asimismo, obtuvimos resultados similares a los de El Ghoul y Karoui (2017) para el mercado de fondos de inversión de Estados Unidos. Utilizando los diferentes pilares de las puntuaciones de ESG (ambiental, social y de gobierno), también alcanzamos una relación negativa entre dichas dimensiones y el rendimiento, demostrando que todas las dimensiones desempeñan un papel importante en la explicación del desempeño financiero.

En términos de riesgo, el nivel de sostenibilidad está relacionado negativa y significativamente con el VaR del fondo, apoyando que los fondos de inversión mejor calificados se comportan mejor frente a las pérdidas extremas del mercado. El signo contrario se encuentra en el caso del parámetro asociado a la variable *dummy* convencional, mostrando las ventajas de emplear una medida cuantitativa de sostenibilidad para evaluar el riesgo de los activos. Este resultado podría significar que los gestores de fondos de inversión SRI basan sus decisiones en un análisis más profundo que resulta en una reducción significativa en el riesgo de sus decisiones de inversión. Nuestra evidencia de trabajo que la puntuación de sostenibilidad puede ser utilizado por los inversores preocupados por las pérdidas extremas y no sólo por los inversores motivados por los valores de sostenibilidad.

Realizamos pruebas de robustez adicionales para verificar la consistencia de nuestros resultados y para proporcionar otros análisis complementarios. Hemos realizado una regresión cuantílica para analizar si los anteriores efectos difieren en los diferentes cuantiles analizados, lo que indicaría diferencias en las habilidades de los gestores respecto el desempeño, no obteniendo diferencias con los resultados previos. Además, se han recalculado los modelos excluyendo del análisis la variable gastos del fondo, dado que dicha variable reducía la muestra, no obteniendo resultados diferentes. Por último, analizando el efecto de la sostenibilidad en los flujos de los fondos se concluye que las rentabilidades no ajustadas al riesgo tienen mayor influencia en las

decisiones de inversión. El score de sostenibilidad afecta de una forma significativa a los flujos de los fondos, por lo que los fondos de mayor calificación tuvieron un mayor volumen de inversión, siendo también significativo el efecto de la variable ficticia de sostenibilidad.

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

Index

ABBREVIATIONS ---- **17**

TABLES ---- **19**

FIGURES ---- **21**

CHAPTER I: INTRODUCTION ---- **23**

1. MOTIVATION OF THE STUDY ---- 25
2. CONTRIBUTION OF THE STUDY ---- 26
3. STRUCTURE OF THE STUDY ---- 27
4. OBJECTIVES ---- 28
5. HYPOTHESES TESTING ---- 28
6. METHODOLOGY ---- 30

CHAPTER II: ARE QUANTITATIVE RATINGS USEFUL TOOLS SELECTING MUTUAL FUNDS? ---- **33**

1. INTRODUCTION ---- 33
2. BACKGROUND MORNINGSTAR STAR RATINGS ---- 34
3. LITERATURE REVIEW ---- 35
4. EMPIRICAL STUDY ---- 41
   4.1. Performance Metrics ---- *41*
   4.2. Models ---- *43*
   4.3. Results for rating models ---- *45*
   4.4. Results with net expenses and other variables ---- *50*
   4.5. Results with downside risk ---- *52*
5. ROBUSTNESS ---- 55
6. CONCLUSIONS ---- 55

CHAPTER III: DOES MORNINGSTAR ANALYST RATING MATTERS FOR MUTUAL FUNDS? ---- **57**

1. INTRODUCTION ---- 57
2. MORNINGSTAR QUALITATIVE RATINGS. ---- 59
   2.1. Stewardship Grade ---- *59*
   2.2. Morningstar Analyst Rating ---- *61*
3. PREVIOUS RESEARCH. ---- 62
4. EMPIRICAL STUDY ---- 69
   4.1. Data and sample ---- *69*

4.2. Models ---- *71*
4.3. Results for Analyst ratings (forward looking) ---- *72*
4.4. Results for some Pillars included in the Analyst ratings (forward looking) ---- *75*
4.5. Results for stars ratings (backward looking) ---- *75*
4.6. Results for best stars and analyst ratings (combining forward and backward looking) ---- *76*
5. ROBUSTNESS ---- 77
6. CONCLUSIONS ---- 80

CHAPTER IV: DOES SUSTAINABILITY SCORE IMPACT MUTUAL FUND PERFORMANCE? ---- 82

1. INTRODUCTION ---- 82
2. LITERATURE REVIEW ---- 86
3. BACKGROUND OF SUSTAINALYTICS´ METHODOLOGY AND MORNINGSTAR ---- 93
3.1 Sustainability Scores. ---- *93*
4. EMPIRICAL STUDY ---- 96
3.1. Sample ---- *96*
3.2. Variables construction ---- *97*
3.3. Performance variables ---- *98*
3.4. Downside risk variables ---- *99*
3.5. Flow of funds ---- *99*
3.6. Descriptive statistics ---- *100*
3.7. Fund Performance and Sustainability Scores ---- *101*
3.8. Downside risk and sustainability scores ---- *104*
3.9. Flows and sustainability scores ---- *106*
5. ROBUSTNESS ---- 107
6. CONCLUSION ---- 111

CHAPTER V. CONCLUSIONS, LIMITATIONS AND FUTURE RESEARCH ---- 113

REFERENCES ---- 118

APENDDIX ---- 127

Abbreviations

CFA- Chartered Financial Analyst

ETFs- Exchanged Traded Funds

ESG -Environmental, Social and Governance

L1- One lag

L3 -Tree lag

MSCI- Morgan Stanley Capital International

MSCI ESG- Morgan Stanley Capital International Environmental, Social and Governance

MRAR-Morningstar Risk-Adjusted Return

NGOs- Non-Governmental Organizations

OLS -Ordinary Least Squares

Q75- Quintile 75%

SR- -Social Responsibility

SRI -Social Responsibility Investment

VaR -Value at Risk

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

Tables

TABLE 1 -STAR RATING.....35
TABLE 2. LITERATURE REVIEW OF PREDICTIVE POWER OF RATINGS.....40
TABLE 3- FUNDS IN THE SAMPLE CLASSIFIED BY STARS (RATING OVERALL).....41
TABLE 4- CONDITIONAL 1 YEAR PERFORMANCE BY RATING AND YEAR.....43
TABLE 5- PANEL DATA WITH 1 LAG (RATING OVERALL).....46
TABLE 6 - PANEL DATA WITH 3 LAGS (RATING OVERALL).....48
TABLE 7- QUANTILE REGRESSION FOR 1 AND 3 LAGS (RATING OVERALL): EQUATION 1.....49
TABLE 8- PANEL DATA WITH ALL VARIABLES BY 1 LAG (OVERALL).....51
TABLE 9- PANEL DATA WITH ALL VARIABLES AND 3 LAGS (OVERALL).....52
TABLE 10.- PANEL DATA WITH ALL VARIABLES AND 3 LAGS (OVERALL).....53
TABLE 11- MONTHLY THREE YEARS CONDITIONAL VAR BY RATING AND YEAR.....54
TABLE 12 - LITERATURE REVIEW OF PREDICTIVE POWER OF STEWARDSHIP GRADES AND MORNINGSTAR ANALYST ON FUTURE MUTUAL FUND PERFORMANCE.....67
TABLE 13.- ANALYST RATINGS DISTRIBUTION FOR THE PERIOD (AUGUST 2012-AUGUST 2016).....69
TABLE 14. ANALYST AND STAR RATINGS DISTRIBUTION.....70
TABLE 15- FUND CHARACTERISTICS BY ANALYST RATING.....70
TABLE 16- CONDITIONAL THREE YEARS PERFORMANCE AND VALUE AT RISK BY RATING.....71
TABLE 17- ANALYST RATINGS AND OUT-OF-SAMPLE PERFORMANCE AFTER 12 AND 36 MONTHS.....73
TABLE 18- PILLARS AND OUT-OF-SAMPLE PERFORMANCE AFTER 12 AND 36 MONTHS.....75
TABLE 19- STARS RATINGS AND OUT-OF-SAMPLE PERFORMANCE AFTER 12 AND 36 MONTHS.....76
TABLE 20- STARS AND ANALYST RATING COMBINATION AND OUT-OF-SAMPLE PERFORMANCE AFTER 12 AND 36 MONTHS.....77
TABLE 21- QUANTILE REGRESSION (OUT OF SAMPLE PERFORMANCE AFTER 12 MONTHS).....78
TABLE 22- ANALYSIS FOR LARGE-SCALE CATEGORIES.....79
TABLE 23.- PREVIOUS RESEARCH ON SRI FUND PERFORMANCE.....91
TABLE 24- PORTFOLIO SUSTAINABILITY SCORE AND SUSTAINABILITY RATING.....95
TABLE 25-PORTFOLIO CONTROVERSY SCORES.....95
TABLE 26- SUSTAINABILITY AND ESG SCORES FOR DIFFERENT GROUPS.....97
TABLE 27- SUSTAINABILITY AND ESG SCORES FOR DIFFERENT GROUPS.....98
TABLE 28- PERFORMANCE METRICS FOR DIFFERENT GROUPS.....98
TABLE 29- VALUE AT RISK FOR DIFFERENT GROUPS.....99
TABLE 30- FLOWS OF FUNDS FOR DIFFERENT GROUPS.....99
TABLE 31- DESCRIPTIVE STATISTICS.....100
TABLE 32- SUSTAINABILITY SCORES AND FUND PERFORMANCE.....103

TABLE 33- ESG PILLARS MODELS AND FUND PERFORMANCE. 104
TABLE 34- SUSTAINABLE SCORE AND DOWNSIDE RISK 105
TABLE 35- SUSTAINABLE SCORE AND FLOW OF FUNDS 107
TABLE 36- QUANTILE REGRESSION. 109
TABLE 37- CROSS SECTIONAL YEARLY ALPHA RATIO 127
TABLE 38- CROSS SECTIONAL YEARLY SHARPE RATIO 128
TABLE 39- CROSS SECTIONAL YEARLY RETURN MODELS 129
TABLE 40- ESTIMATES USING THREE YEARS RATING INSTEAD OVERALL. 130

Figures

FIGURE 1- SUSTAINALYTICS FRAMEWORK. ---------- 93

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

Chapter I: Introduction

This thesis is a research development that aims to respond some of the various issues that may arise for mutual fund investors. An important question in fund´s investment is the selection of funds to be integrated in the portfolio. Ratings are used by most of investors to select mutual funds, so there are several studies as Faff, Parwada and Poh (2007), Del Guercio and Tkac (2008) that support the evidence that investors take their decisions based on ratings.

More specifically, this thesis aims to deepen the knowledge and the role of Morningstar's rating, that is used to make investment decisions in the selection of mutual funds. Morningstar is the market leader in providing independent research to financial advisers such as individual and institutional investors (banks, funds, insurance companies, etc.).

It stands out from other institutions because it does not depend financially or have any direct relationship with the evaluated investments, unlike other rating agencies. To elaborate their scores and ratings Morningstar developed a strategy of grouping mutual funds with similar categories, dividing them according to their characteristics into different categories. Morningstar came to realize that the prospect of mutual funds did not explain effectively the fund's investment strategies. For instance, many funds claimed to be seeking "growth" by investing in some top-line companies, while others were investing in Small Business Enterprises but Morningstar scores and ratings have solved this problem by grouping portfolios into peer groups based on their holdings.

It seems consensual that investors care more about Morningstar's quantitative ratings, in fact, studies such as Faff, Parwada and Poh (2007) and Del Guercio and Tkac (2008), among others, conclude in their studies that funds have better quantitative or Star Ratings (which are based on past risk-adjusted performance measure and is rated from one to five stars like the hotels), are able to increase the inflows of funds and the worst ratings cause outflows. Best funds are rated with five stars, and the worst with one star.

There is a classic debate about whether past returns may indicate the ability to achieve future returns and that has to do with the manager's ability to repeat past performance, therefore the matter relies on the doubt of weather it makes sense to base investment decisions on these ratings.

In fact, it will only makes sense if past returns can have predictive power of future returns. Authors such as Blake and Morey (2000) and Morey (2002, 2005) among others, studied the predictive power of quantitative ratings obtaining weak results. However, some recent studies have found that ratings may lead to the choice of the best funds and ratings can predict, in

some extent, future performances. Authors such as Morey and Gotesman (2006), Müller and Weber (2014) and Meinhardt (2014) have demonstrated through their studies that quantitative ratings may be able to predict the future performance of mutual funds.

Star ratings are the most popular among mutual fund investors, because ratings are goals based on historical performance and easy to understand. However, Morningstar has also developed other ratings that further enhance the quality aspect and can be useful, and even complement quantitative ratings, by helping to understand whether ratings can contribute to the selection of the best funds. Among the most popular are the Star Ratings or quantitative ratings, but also the qualitative Analyst ratings and the Stewardship Grades, including the recent Morningstar's Sustainability Rating.

The selection of funds based exclusively on historical performance or quantitative ratings, excludes a set of qualitative factors that can explain future performance. This is why, in addition to quantitative ratings it has appeared that ratings based on analyst opinions evaluate aspects of mutual funds as: Governance, Process, People, Parent, Board Quality, Corporate Culture, Fees, Manager Incentives or Regulatory Issues. The existence of several studies about quantitative and qualitative ratings are not as popular since there are only a very limited number of researches that focus on the the ability to select good funds based on qualitative ratings.

In the particular case of Morningstar, there are two alternatives: Analyst Ratings and Stewardship Grade. In our study of qualitative ratings, we will focus on Analyst ratings as it is the less discussed subject in literature, since there are more recent studies like Morningstar who launched them in 2011.

Analysts ratings rate the funds based on their belief in the fund's ability to outperform its benchmark or its competitors in the long run. To reach a rating, analysts evaluate five key pillars that our experience has shown as critical to a fund's ability to succeed like: Personal, Management Firm, Process, Performance, and Price. In relation to the star ratings there are some studies trying to understand the role of these in the selection of mutual funds. Regarding qualitative ratings, there are not many studies, since they are relatively recent.

Wellman and Zhou (2007), Seng (2009) and Gottesmann and Morey (2012) studied the Stewardship Grades, and Kamal (2013) and Armstrong, Genc y Verbeek (2016) are the unique authors that studied the predictive power of Analyst ratings. In general, their results have shown the ability to predict future out-of-sample performance: higher Analyst Ratings do predict better future performance and it is verified in funds with better rating (Gold medal).

Finally, we investigate the effect of Morningstar's Sustainability and ESG scores on performance, flows and risk. These scores are compiled for each fund that invests at least 50% of its assets in companies rated by Sustainalytics (a leading company in the ESG criteria valuations

of companies worldwide). It's methodology consists of different ESG indicators between 120 and 150, that depending on the industry, measure the sustainable the companies practices.

Sustainalytics evaluates the company's performance in each of these indicators based on several internal and external data sources (Sustainalytics, 2016), it uses two types of indicator templates: basic and sector-specific. Each company obtains a score by aggregating the indicators for each of the ESG pillars, the global ESG score and the controversies of the company.

Morningstar adds the asset scores that form the portfolio of an investment fund in a standardized way to obtain the total score of the fund in each ESG pillar, total ESG score, fund controversies, and finally, by difference between the score ESG and the controversies, the sustainability score of an investment fund.

Although according to Lewis and Mackenzie (2000) and Webley, Lewis and Mackenzie (2001) some investors in socially responsible investment (SRI) funds are willing to accept lower returns for their moral stance making the performance of SRI funds and conventional funds an open question. As Junkus and Berry (2015) sustain, after a review of the most recent work in major finance journals on SRI: "the performance of SR mutual funds and indexes are not generally significantly different to conventional funds or indexes, but again these results are also highly dependent on model specification, time period, benchmark, and other characteristics of the study".

As far as we know, only El Ghoul and Karoui (2017) use a scores to study the effect on sustainability on fund performance and flows, concluding that higher values displaying poorer performances and weaker performance-flow relations.

1. Motivation of the study

The motivation behind this study lies in the diffusion of the use of funds´ ratings in the making of their investment decisions. The Morningstar´s ratings are the most important in the industry of mutual funds because it is a very respected investment research firm among investors, sustained by many studies to prove that investors care about the Ratings such as: Faff, Parwada and Poh (2007) and Del Guercio and Tkac (2008), among others that find quantitative ratings important for the flows of mutual funds. This raises the question of whether investors are well advised to pay attention to Morningstar ratings.

There are many funds, with different strategies and different assets so it is impossible to know and control all the information about a huge amount of funds and assets with many specification. Only the specialists of the various branches of investment have enough knowledge to do it, and it is impossible for non-institutional investor to know all the investment strategies and understand them.

Persistence plays a very important role on the literature debate that allows us to see if there is the ability of certain managers to achieve a better performance by beating the market in a consistent way. So, it will be important to see if investors can make their investment decisions and to realize if extent past performance may contain information about future performance. The question whether qualitative aspects in management are more determinant or can complement the quantitative information, as is the case of historical returns, will be another important aspect to realize in this thesis.

Finally, this work is also of interest, because Morningstar is following the needs of investors and in the most current debates. An example of this is not only the qualitative aspects that were a clear response to Morningstar's new investor needs but also the creation of a rating that evaluates SRI behaviors of mutual funds so it is an important motivation to realize to what extent SRI affects performance. Finally, choosing the best funds with the information contained in the ratings can be a very efficient and simple way for investors, since the ratings greatly simplify all the information contained in the investments

.

2. Contribution of the study

This study intends to give several contributions to the academic literature, as well as to help in the decision making of the investors and institutions. There are many studies that focus on the quantitative ratings of mutual funds and the persistence of returns to see if there may be information on past returns to predict future returns. However, there are not many known studies that focus on the predictive power of Morningstar Star Ratings.

Some studies such as Blake and Morey (2000), Morey (2002, 2005) and some others, studied the predictive power of quantitative ratings obtaining weak results, but some recent studies have found that ratings may lead to the choice of the best funds that ratings can predict, in some extent, future performances.

Authors such as Morey and Gotesman (2006), Müller and Weber (2014) and Meinhardt (2014) have demonstrated through their studies that ratings may be able to predict the future performance of mutual funds, but none of them used a database with so many funds and with such a long-time spectrum like as ours.

There are few studies that have focused on the predictive power of star ratings with databases that have a significant share of funds with ratings but after the change in methodology we do not know any of them. In this empirical analysis, control variables such as Net Expenses, Turnover, among others, are also added to see if they can help to increase the predictive power of

star rating when used together, or to also understand if these control variables can explain the star ratings and past performance.

In the following empirical analysis, we focus on Analyst ratings. Kamal (2013) and Armstrong, Genc y Verbeek (2016) are the unique authors that studied the predictive power of Analyst ratings, but no one has studied Analyst ratings and their predictive power in an exhaustive way. The existing studies use a database with few funds and the time horizon is very short, since Morningstar's analyst ratings are relatively recent.

In this study we develop a rating in which we mix the star ratings and qualitative to see if we can combine the qualitative and quantitative ratings to select the best funds. Control variables are also added to see the explanatory effect in relation to ratings.

In the last empirical study, an innovative study is carried out, where the debate on the outperform, under-perform and neutral effect of sustainable investments is taken into account. Although there are some studies on this effect, no one has used these new ratings from Morningstar (they are fairly recent) and studied their effect on future performance of mutual funds. On the other hand, existing studies on known sustainable investments use only one dichotomous variable, which makes it impossible to perceive in what level, sustainable investments are beneficial, neutral or harmful to create value for mutual funds.

3. Structure of the study

This thesis is divided into five chapters.

Chapter I consists in a description of what it is proposed in this study: the reason behind our interest, it's contribution to the knowledge of this topic, the goals to achieve, making a description the hypothesis and the methodology used.

Chapter II is the first empirical study that is based on the predictive power of Morningstar's star ratings divided into: introduction, literature review, a background of ratings and performance measures, methodology adopted, empirical test and conclusions.

Chapter III is an empirical study about investments predictive power of qualitative ratings (analyst ratings). A study is made about the predictive power of Morningstar's analyst ratings in the mutual fund's performance. The structure is as follows: introduction; literature review; a background of ratings and performance measures; methodology adopted; empirical test and conclusions.

Chapter IV it is another empirical study related to the sustainability scores of the portfolio of a mutual fund and its performance. We study the effect of investments socially responsible using the Morningstar's Sustainability and ESG Scores followed by this structure: introduction;

literature review; a background of ratings and performance measures; methodology adopted; empirical test and conclusions.

The last one, Chapter V, is summary of the conclusions drawn from the three empirical studies, as well as all conclusions drawn and future research.

4. Objectives

Many investors try to diversify their portfolios by investing through mutual funds, so their selection of funds that will integrate a portfolio is an important question. Since ratings are used by most of investors to select mutual funds, the main objective of this thesis is to analyze the usefulness of Stars ratings, Analyst ratings and Sustainability and ESG Scores of Morningstar in fund selection**.** The use of the Morningstar ratings is because it is the most important agency in the mutual fund industry.

This broad objective can be broken down into different specific objectives:

- To analyze if Morningstar's quantitative ratings (Star ratings) can explain future performance and downside risk.
- To investigate if Morningstar's Analyst ratings predict future performance.
- To investigate if the combination of the qualitative and quantitative aspects of the Morningstar ratings can predict future performance.
- To analyze if mutual funds that invest in SRI have superior, inferior or neutral performance compared to conventional investments.

5. Hypotheses Testing

In Chapter II we test the following hypotheses:

- H1: Mutual funds with better Star ratings will have better performance in risk-adjusted returns.
- H2: Mutual funds with better Star ratings will have lower Value at risk (VaR).

In Chapter III we test the following hypotheses:

- H3: Mutual funds with better Analyst ratings will have better performance in risk-adjusted returns.
- H4: Mutual funds with better Analyst ratings will lower Value at risk (VaR).
- H5: Mutual funds with better Analyst ratings and better Star ratings will have better performance in risk-adjusted returns.

• H6: Mutual funds with better Analyst ratings and better Star ratings will have lower Value at risk (VaR).

In Chapter IV we test the following hypotheses:

• H7: mutual funds with better ESG and Sustainability Score have neutral performance compared.

• H8: mutual funds with better ESG and Sustainability Score have lower Value at risk (VaR).

6. Methodology

The thesis is divided into five chapters: Chapter II "Are quantitative ratings useful tools selecting mutual funds?"; Chapter III "Does Morningstar Analyst Ratings Matters for Mutual Funds?"and Chapter IV: "Does sustainability Score impact in performance?".

Throughout this work, the statistical program used was the statistic program Stata 12 and the database used in all empirical studies was Morningstar Direct ™.

In the second Chapter, we use the analysis of Stars Rating of Morningstar.

The first method we use to examine the out-of-sample predictive performance is panel data regression model (Pooled and Random effects). The methodology based on panel data can control individual effects with advantages like the reduction of colinearity and efficiency, among others (Baltagi, 2010). We also use this methodology with 1 and 3 year lags, in order to evaluate the predictive ability of Morningstar ratings for medium and short-term performance. For each term, we use specific measures and data to avoid autocorrelation problems.

First, we estimate the models using exclusively ratings to assess the effect of using only this variable for funds selection, later, we include other control variables, like expenses, size or manager experience.

Instead of Panel data and based on Chen and Huang (2011), we also use the Quantile Regression to extend the regression model to conditional quantiles of the different performance metrics because it is more appropriate for a heterogeneous mutual fund universe, where strategies and objectives can vary. This model let us capture information about the coefficients at different quantiles of the dependent variable given the set of endogenous variables (star rating). In addition, the conditional quantile regression developed by Koenker and Bassett (1978) deals well with skewed distributions of fund performance. In particular, we adopt the bootstrapping method proposed by Efron (1979).

In the third Chapter, we analyze if the Analyst rating can help to identify products that will outperform their peers in the following after the initial rating. At the same time, we also include separately stars ratings to compare both alternatives, taking into account that investors take their decisions based on backward looking data base. Finally, we use both ratings to evaluate if taking decisions combining good stars and analyst mutual funds, can help in the selection of out performers. Firstly, we estimate some models to evaluate the sample performance based exclusively on the analyst rating and their main pillars. If analyst rating is a forward-looking measure that reflects the expectations of analyst about future performance in the long run, we expect that higher ratings will obtain higher future performance. We calculate out

of sample risk adjusted returns to 12 and 36 months after the initiation rating and then we regress to the different metrics using Gold, Silver and Bronze indicators variables.

For 12 months models, we estimate a panel data regression model (random effects) and for 36 we use ordinary least squares (OLS) regression because can we only have one period of three years after the initial grade. The methodology based on panel data can control for individual effects with advantages like the reduction of colinearity and efficiency. In addition, we also use the Quantile Regression to extend the regression model to conditional quantiles of the different performance metrics.

In the fourth Chapter, we use OLS regression that we only have sustainability data on November 2016. We analysed the performance and the risk of the mutual funds with ESG and Sustainability Scores data of the last two years based on the work of Wimmer (2012), which shows that the portfolios of the funds regarding sustainability do not vary to that term (similar to the previous chapters).

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

Chapter II: Are quantitative ratings useful tools selecting mutual funds?

1. Introduction

Mutual funds are one of todays' leading saving products. An important question which investors have to face is the selection of funds to construct their portfolios. Ratings are used by most of the investors to select mutual funds; in fact, many investors make their choices based exclusively on these scores. Some studies such as Faff, Parwada and Poh (2007) and Del Guercio and Tkac (2008), support the evidence that investors make their decisions based on ratings. Their results showed that the variations in the flows of mutual funds are affected specially by the change in ratings. Despite the fact much research has been done in the field of mutual funds, in the particular case of ratings there are few studies and their results are not conclusive.

One important question to answer is the relative to the ability of ratings to predict the future performance of mutual funds. In this sense, the main question we want to assess in this study is if Morningstar ratings are reliable tools selecting funds. Authors such as Blake and Morey (2000) and Morey (2002, 2005) among others, studied the predictive power of quantitative ratings obtaining weak results. However, some recent studies have found that ratings may lead to the choice of the best funds, and ratings can predict, in some extent, future performances. Authors such as Morey and Gotesman (2006), Müller and Weber (2014) and Meinhardt (2014) have demonstrated through their studies that ratings may be able to predict the future performance of mutual funds.

In this paper, we evaluate the ability of quantitative ratings to anticipate the future behaviour of the fund's performance. In addition, is important to check what is considered as a good rating and if there are performance differences between all categories or only between bad and good ones. We also evaluate the ability to anticipate short and long-term funds´ performance. In addition, we analyse if ratings also have differences in terms of downside risk through the analysis of VaR (value at risk). Finally, we include the interaction with costs and other variables such as size, management quality or categories, among others.

Our results support the ability of quantitative ratings to predict future performance in the short and medium term even after the inclusion of management fees. Moreover, the best ratings have a better behaviour in terms of VaR helping to preserve the investors´ wealth. This work is important to the academic world, but also it is interesting for investors, rating companies and financial institutions.

The paper is organized as follows. First, we describe Morningstar Star Ratings methodology. Then, we review and summarise the main existing research about quantitative ratings. Finally, we present the empirical analysis, the statistical model and we show the contrast between the assumptions and the main results.

2. Background Morningstar Star Ratings

Morningstar is an investment research and management company, which provides data of mutual funds, among other investment offerings. The original Morningstar ratings were introduced in 1985 and the challenge was to design a rating system to help investors in the funds selection. Recently, Morningstar has extended their services to another class of assets. The methodology of Star Ratings evaluates funds around the world based on historical data. The Stars run from 1 star (lowest) to 5 stars (highest) and is characterized for basing their ratings on fund`s risk-adjusted returns. The first step is to classify funds in Morningstar Categories. Morningstar use Category Peer Groups to make distinction funds. These groups of funds define funds that can have similar orientation by value-growth as industry sector; Beta; geographic region and country weights; duration and credit quality; historical return volatility; among others investment style factors. Morningstar uses a scale of one to five Stars like hotels and it is an effortless way to interpret the quality for investors. In 2002 Morningstar increased and changed groups and created a new way to measure risk-adjusted returns. Other important changes added in 2002 were the analysis of Mutual Funds with one Portfolio and multiple shares classes and also their expenses structure.

Now, Star Ratings are based on risk-adjusted returns and all costs as sales charges, loads, and redemption fees are included for 3, 5 and 10 year periods. Funds with less of 3 years are not rated. Morningstar Risk-Adjusted Return (MRAR) is obtained through the utility function, which is based on expected utility theory. The utility function is based on the criteria that funds with better return and same risk are more preferable. Ratings include all variations in monthly returns and also given more emphasis by consistent performance and reduce the probably of strong short-term performance masking inherent fund risk. Morningstar Ratings are rated as can be shown in Table 1. As you can see the 10% highest funds risk-adjusted return are rated as 5 Stars, 22.5% rated 4 Stars (above-average performance); 35% rated 3 Stars (average performance); 22.5% rated 2 Stars (below average performance); and 10% rated 1 Star (poor performance). Moreover, Star Ratings have rated by periods of 3, 5 and 10 years and Rating Overall that is a weighted average of the 3-year, 5-year and 10-year ratings with weights depending on the age of the fund and rankings are revisable each month.

Table 1 -Star Rating.

Fund percentile	Rating
Top 10%	*****
10% to 32.5%	****
32.5% to 67.5%	***
67.5% to 90%	**
Bottom 10%	*

3. Literature Review

Does past performance predict future performance? That`s an age-old question in literature and a lot of authors find support for this theory and academic researches named it as persistence. The persistence in mutual funds has generated much interest in the research community. Persistence is the ability of a particular fund to repeat the same or similar future performance rather than past performance. Persistence support the ability of mutual fund managers to produce losers and winner's funds. A lot of studies find evidence of persistence in funds returns (Hendricks, Patel and Zeckhauser,1993; Brown and Goetzmann, 1995; Grinblatt, Titman and Wermers, 1995; Carhart, 1997; Wermers,1999; Grinblatt and Keloharju, 2000; Carhart, Carpenter, Lynch and Musto (2002) Nofsinger and Sias, 1999; Bollen and Busse, 2005; Vidal-García, 2013, among others). Hendricks, Patel and Zeckhauser (1993) concluded that of no-load, growth oriented mutual funds persists in short-term, specifically, the evidence is stronger for a one year evaluation horizon. The poorer portfolios do significantly worse than the standard benchmarks and those of top performers mutual funds behave better than the benchmark. The difference between the top and the bottom octile is six to eight percent per year. Brown and Goetzmann (1995) explore performance persistence in mutual funds using an absolute and relative benchmark. The study indicates that performance adjusted risk persists, however, performance is mostly due to funds that the S&P 500. Also, that the poor performances persist and increase the probability of disappearance of the funds. Grinblatt, Titman and Wermers (1995) analysed the extent to which mutual funds purchase stocks based on their past returns. Authors find that 77 percent of mutual funds were "momentum investors" buying stocks that were past winners. On the other hand, on average, funds that were invested on momentum achieved significantly better performance than other funds. Carhart (1997) find that common factors in stock returns and investment expenses can explain the persistence in equity mutual funds and "hot hands" results mostly driven by one-year momentum effect. Wermers (1999) analyse the trading of mutual funds and the "herd" effect on stock prices. It concluded in the study that stocks although it found a little herding in mutual funds, it finds a strong herding on trades and stock prices. Stock that herds buy outperforms stocks that it sells by 4 percent during the following six months. Nofsinger and Sias (1999) argue a strong and positive correlation

between changes in institutional ownership and returns measured over the same period. Authors find that evidence that both factors can explain the connection between herding prices and the herding by individual investors and herding is positively correlated with lag returns. Grinblatt and Keloharju (2000) study from Finland if past returns determine the propensity to buy and sell and whether these differences in past-return-based behaviour. Authors find that foreign investors, in general, buy winners and sell losers funds, contrarily particularly households' investors. The portfolios of foreign investors outperform particularly households' investor. Carhart, Carpenter, Lynch and Musto (2002) provided a study about a mutual funds survivorship and their relationship with performance persistence. Authors find that conditioning on survival weakens the evidence of persistence and showed how the relationship between performance and fund characteristics can be affected when it uses a survivorship sample. The magnitude of biases in mutual funds can be explained for fund size, expenses, turnover, and load fees. The authors concluded that there exists a strong persistence in mutual funds, if the sample was free of survivorship. Bollen and Busse (2005) estimate parameters of standard stock selection and market timing models and it ranked stocks of the average abnormal return of the top decile. The post-ranking abnormal return disappears when funds are evaluated over longer periods. Vidal-García (2013) examines the performance and persistence in performance of European equity mutual funds between 1988 and 2010. The authors documented strong evidence of persistence in benchmark-adjusted returns over 1 year on time horizons of up to 36 months. However, the persistence is much pronounced for the top and bottom performers. It can be concluded that European mutual funds have explanatory power for future performance.

Despite being a less discussed issue, persistence, is also an object of study in relation to ratings, that is, the ability of mutual funds for getting consistent returns above average. Authors such as Howe and Pope (1996), Blake and Morey (2000), Morey (2005), Morey and Gottesman (2006) Duret *et al.* (2008), Philips and Kinniry (2010) and Chotivetthamrong (2015) concluded that ratings have little power predicting future performance. In particular, Howe and Pope (1996) analyse Forbes ratings of stock funds in an attempt to verify if they have the power to predict future performance. They concluded that Forbes up-market ratings can predict future Betas; however, they have weak power predicting future performance, particularly in risk-adjusted returns. Furthermore, the authors found that the Forbes down-market ratings can help to predict future performance returns over the following year and perhaps even for the entire sixteen-year period study. However, it is not so clear whether this is true predictive power ability or if it is a spurious relationship. In the same line, Blake and Morey (2000) compare the marks awarded by Morningstar with a methodology based on historical average rates, which is called "naive predictor". The lower ratings of Morningstar usually indicate that the future performance will be

weak. In most cases, there is no statistical evidence that funds with good Morningstar ratings have a better future performance than the average of other funds. The research concludes also that the Morningstar ratings are not better at predicting performance than the "naive predictor" methodology. In addition, they found weak statistic evidence for higher yields of 5-star ratings over 3-star or 4-stars ones. However, they observe that lower ratings keep weak performance in the future, for future periods of three and five years, respectively.

Morey (2005) focused their study on mutual funds after reaching the 5-star rating, concluding that the fund´s profitability will fall in the next 3 years after obtaining this score. The author finds that when a fund reaches the 5-star level the risk increases and the managers have difficulties managing capital inflows. This study advises investors to be cautious when a fund has a higher rating, since there are strong possibilities of lower profitability in the next three years. Authors such as Duret *et al.* (2008) studied the ability to predict future performance through external ratings. Many investors use mutual funds with the 5 Star rating to build their investment portfolio. The implicit idea of these authors is as follows: funds that were better in the last three years will have better performance than the others in the future. To develop this study the authors, use the Markov methodology based on Garnier and Pujol (2007). The authors demonstrate that there is weak persistence, which means that the choice of mutual funds with the 5-star rating cannot predict the best future performance. Supporting these achievements, Philips and Kinniry (2010) also looked at Morningstar mutual fund ratings and the future performance and tried to see if the funds with higher ratings exceed those that have lower ratings in the future. In this research, the authors conclude that there is an extreme difficulty predicting future performance. The ratings make a great effort to explain past performance but in fact they have little information about future performance. The authors also criticize investors who use the star ratings as the only criterion for fund´s selection. Investor must not select funds based exclusively on ratings and they ought to consider other qualitative aspects such as the parent organization, the managers, maintenance costs and commissions. These combined aspects can achieve better long-term results for investors. These authors also concluded that funds with higher costs have lower returns than mutual funds with lower expenses, suggesting indexing as a powerful strategy to achieve good results. Chotivetthamrong (2015) monitors during 5 years (2003 to 2007) the funds' Morningstar rating and the performance in the Thailand market. She separates the funds into three groups (above-average funds, average funds or 3-stars Morningstar funds, and below-average funds) finding that most of above-average and average funds dropped their performance within 5 years; however, the funds below average increased or stayed stable their performance.

However other authors such as Ferson and Schadt (1996), Morey and Gottesman (2006), Brenning and Fritzén (2009) Antypas, Caporale, Kourogenis and Pittis (2009), Müller and

Weber (2014), and Meinhardt (2014) concluded that the ratings may be useful predicting future performance. The authors used the framework established by Ferson and Schadt (1996) concluding that higher ratings of Morningstar are associated with higher future returns. In the study, they take into account all the star ratings and they find that funds with 5, 4 and 3 stars have better results than the benchmark. On the other hand, it is clear that the lower ratings (star1 and star2) identify the worst funds easier than the higher ratings the bests. In particular, the authors are interested in examining how the highest ratings reflect superior skills in managing mutual funds. In this study, it was demonstrated that higher ratings are associated with higher yields. On the other hand, it also seems clear that lowest ratings achieve lower performances.

In fact, Morey and Gottesman (2006) find that the new methodology of Morningstar (year 2002) is able to predict future performance of mutual funds. The authors use the new methodology and they analyse 3,886 mutual funds from July 2002 to June 2005. In their study, they find predictive power for the next three years out of sample. Another important conclusion is that higher rated funds significantly outperform lower rated funds and that lower rated funds with 2 stars significantly outperform then lowest rated funds of 1 star. In turn Antypas, *et al.* (2009) evaluated the quantitative Morningstar ratings. The authors applied a simple strategy to probe if higher rated funds reflect superior performance or not. This strategy consists of selecting the highest-rated funds when creating a portfolio. The authors examine the statistical properties of the five funds-of-funds return series and the returns including Star5 to Star1. The results indicate that only the three highest-rated categories of funds are characterized by some funds selection ability, whilst none of the five categories exhibit market timing ability. Overall Star Rating system are the most effective in identifying the worst-performing funds than the best-performing. Also, Brenning and Fritzén (2009) studied if changes in Morningstar Ratings affect the performance and the investment strategy of mutual funds. They use a sample of 223 shares of mutual funds between January 2002 and February 2009, taking into account Morningstar rating changes in the following two years. They found a significant change in net flows of mutual funds when Morningstar ratings changes, but it is not evident the effect on performance and the investment strategy. The authors concluded that in the following year the upgrade there was an improvement in the fund's performance; in turn, when funds suffer a downgrade in the Morningstar rating it is not so clear that the mutual fund suffers from a lower performance the following year. The authors also concluded that mutual funds managers reduce the risk of their portfolios the following year after receiving a high rating. In turn; when there is a downgrade of the rating the effect on the fund management is more ambiguous. In the same line, Müller and Weber (2014) tested the ratings systems of Stiftung Warentest (protection agency consumers and a major provider of fund ratings in Germany) to see if the ratings of this agency have the power

to predict future performance in the German investment fund market, where they shared the funds for six categories: German, Eurozone, North America, Pacific and World. The authors concluded that there is a significant relationship between the ratings and its future performance, but this relationship is not the same for all the categories of funds. For example, there is statistically significant evidence of World category bond funds, but there is no such evidence for the US bond funds. The ability to predict future performance depends on several factors such as the measurement of performance, the investment horizon and the evolution of the technique used. In turn, the persistence of performance is not evident for all types of funds; it is clear in some cases and not in others. Also in the German market Meinhardt (2014) studied the ability to predict performances by three German rating agencies: Feri Trust, Finanztest and FondsNote. The study focuses on whether these three agencies have the power to predict performance for German investment funds, where it has been shown that when it takes into account more than one agency to predict future performance in mutual funds, the predictive power increases substantially. Table 2 summarizes the main studies reviewed of predictive power of ratings.

Table 2. Literature review of predictive power of ratings

Authors	Sample	Number	Predictive Power	Conclusions
Blake and Morey (2000)	*Two samples: Seasoned funds 1992-1997 and complete funds 1993*	*Morningstar On-Disk and Principia disks. U.S. domestic equity funds with Morningstar rating. 1993 sample group (635 funds).*	*No*	- Is relatively easy to predict poor performance, but it is much more difficult to predict superior performance. *- Low ratings indicate a poor performance.* *- "Naïve predictor" has more predictive power than 4 and 5 Stars.*
Morey (2002)	*September 1991 to September 2000*	*Morningstar On-disk or Principia. Morningstar's Domestic Equity Category Depending on the year, ranging to less than 1,000 in 1991 to near 4,000 in 2000.*	*No*	*-Average overall star ratings of seasoned funds are consistently, and sometimes significantly, higher than the younger funds.* *- Methodology Morningstar is biased.* *- The weight and rounding used by Morningstar decline in overall ratings relatively more difficult for seasoned funds.* *- Funds ratings haven`t much ability to predict future performance*
Morey (2005)	*July 1993 to July 2001*	*All funds of Data Disk with rated a 5-star (273 funds)*	*No*	*- Expenses, Portfolio and Turn Over do not change much after received 5-Star Rating.* *-Three years after a fund received its initial 5-star rating, fund performance falls of.* *- Morningstar Ratings itself having a fantastic influence in fund flows.*
Morey and Gottesman (2006)	*July 2002 to June 2005*	*3,886 funds of Morningstar Principia mutual funds data of domestic equity funds*	*Yes*	*-The study tries to understand how the new Morningstar rating System, predicts future performance* *- Authors find that new system Morningstar Ratings can predict future performance in three years out the sample* *- Higher rated funds significantly outperform lower rated funds* *- The next to lowest rated funds (2- star funds) significantly outperform the lowest rated funds (1-Star funds)*
Del Guercio and Tkac (2008)	*November 1996 to October 1999 Morningstar s Principia*	*3,388 domestics equity mutual funds – Morningstar, Inc.*		*-Authors argue that only 5-Star ratings can attract new assets* *- Excluding the rating initiation categories, flows change with downgrade and upgrade of Star Ratings*
Duret et al. (2008)	*(*)*	*(*)*	*No*	*- Using a Markov modelling authors show that ratings persistence is poor*
Antypas et al. (2009)	*January 1998 to "paper date"*	*Morningstar Direct Data 1,511 Equity funds quoted in US Dollars*	*Yes*	*- Better performances in Star3, Star 4, Star5 reflects stock selection rather than market timing abilities.* *- It´s more common that the worst funds with Star1, Star2 will also be worst in the future. So, Morningstar ranking system is most effective in identifying the worst-performing funds rather than the best-performing ones.*
Füss et al. (2010)	*May 2004 to April 2009*	*2,490 funds of Morningstar Inc.*	*No*	*- Results suggest that Morningstar Ratings have a little ability predicting future performance.* *- It`s possible to identify funds that will have poor performance in the future.* *- One-star ratings have worse results than the five-star category.* *- Morningstar rating has a big correlation with three-year Sharpe ratio.* *- Sharpe-ratio can be better at predicting future performance than Morningstar ratings*
Müller and Weber (2014)	*December 2001 to June 2008*	*Stiftung Warentest fund rating system (2,351 mutual funds for 30.06.2008)*	*Yes*	*- Past performance is positive related with future performance, in many funds.* *- This study reveals significant differences in persistence between different funds categories.* *- Predictive power depends on active or passive strategies.*
Chotivrtthamrong (2015)	*2003-2007*	*36 Thai mutual funds form Morningstar*	*Yes*	*- Paper suggests a positive relationship between Morningstar Rating and performance.*

(*) Methodological paper.

4. Empirical study

Our sample contains 1,579 European equity funds rated by Morningstar from 2004 to 2014. We limit our sample to the funds included in the following Morningstar categories: Europe Flex-Cap Equity, Europe Large-Cap Blend Equity, Europe Large-Cap Growth Equity, Europe Large-Cap Value Equity, Europe Mid-Cap Equity, Europe Small-Cap Equity, Europe ex-UK Large-Cap Equity, Europe ex-UK Small/Mid-Cap Equity, Eurozone Flex-Cap Equity, Eurozone Large-Cap Equity, Eurozone Mid-Cap Equity and Eurozone Small-Cap Equity. The sample contains 10,375 observations, where 6.89 % have 1 Star, 19.37% have 2 Stars, 35.89% with 3 Stars, 26.46 % with 4 Stars and finally 11.38 % with 5 Stars. The funds are the type "open funds" with quantitative overall rating Morningstar in the investment area of total Europe. Investment area identifies the geographic area that the fund focuses its investments in. Funds selection has been done for all funds independently they were active or not to avoid survival bias. Furthermore, to avoid problems of multicollinearity, we select only a class for each fund, choosing these sequence preferences: institutional class, lower management fee, the lower net expense ratio, higher class size, oldest start date and accumulation preferred. In Table 3 the distribution of the rating is showed.

Table 3- Funds in the sample classified by Stars (Rating overall)

Rating Overall	Freq.	Percent
1 Star	760	7.33%
2 Stars	2,169	20.91%
3 Stars	3,749	36.13%
4 Stars	2,575	24.82%
5 Stars	1,122	10.81%
Total	10,375	100%

4.1. Performance Metrics

To measure out-of-sample performance we use two risk-adjusted metrics: Alpha and Sharpe ratio. In addition, we include a non-adjusted one (annual return of the mutual fund) and the Value at Risk (99%) as a measure of downside risk. We will now explain briefly the -performance metrics:

a) Sharpe Ratio is a way of measuring the expected return per unit of risk for a zero-investment strategy. Since it was introduced by William Sharpe in the 1960s, the Sharpe ratio has become one of the most widely used metrics in finance and economics. The Sharpe ratio is one of the tools most widely used metrics in finance and economics that help investors evaluate the relationship

between risk and return of asset. By quantifying both volatility and performance, this tool allows for an incremental understanding of the use of risk to generate return.

The Sharpe ratio can be represented by:

$$S_{it} = \frac{(R_{it} - R_f)}{\delta}$$

Where:

- Sit is adjusted-risk return
- R it represent the fund return
- $R\ f$ represent the risk free
- **δ** is the standard deviation

b) We can consider Single-index alpha a return measure risk-adjusted, widely used to measure the returns of mutual funds. In fact, Alpha measures the return investment relative to the benchmark in which the assets are located. That means that single-index model is a simple way to measure a return of a portfolio, that it is represented by:

$$R_{it} - R_f = \alpha_i + \beta_i (R_{mt} - R_f) + \varepsilon_{it}$$

Where:

- $R_{it} - R_f$ is the excess return on the market
- α_i is the abnormal return
- β_i is the stocks' beta
- ε_I are the residual (random) returns

c) Annual Return is a metric to measure returns of Portfolio where the value is comparing with the time horizon previous.

$$AR = \frac{Rit - Rit-1}{Rit}$$

Where:

- R_{it}=Value of Portfolio
- R_{it-1}= Value of Portfolio in t-1

Table 4 contains one year performance ahead conditioned to the previous rating. We can observe that generally highest-rated funds outperform persistently the worst rating in

terms of Sharpe, and in most of cases in terms of yearly return and Alpha, with the exception of 2008 and 2009.

Stars	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014
	Alpha										
1 star	-2.73	-2.21	-0.80	-0.32	30.18	30.56	10.40	-5.20	-1.34	-1.10	-1.58
2 star	-3.41	-1.00	-2.32	00.51	10.57	10.89	-0.14	-2.97	10.37	10.38	-3.55
3 star	-1.90	-1.47	-2.54	00.10	10.78	10.61	00.93	-2.41	10.09	10.47	-1.87
4 star	-0.03	00.01	-2.87	00.10	20.27	30.44	10.49	-1.77	10.99	20.94	-1.31
5 star	-0.59	-0.98	-1.24	-0.57	10.64	10.63	30.42	-2.48	40.36	40.38	-0.46
	Sharpe										
1 star	-0.46	-1.04	10.00	10.17	0.84	-0.71	-0.54	0.13	-0.12	0.13	0.89
2 star	-0.39	-0.94	10.12	10.50	10.12	-0.63	-0.44	0.32	00.12	00.29	10.08
3 star	-0.23	-0.88	10.24	10.73	10.13	-0.58	-0.38	0.39	0.25	00.44	10.32
4 star	-0.03	-0.80	10.52	10.85	10.19	-0.52	-0.30	0.49	00.38	00.56	10.46
5 star	0.29	-0.74	10.86	20.10	10.26	-0.43	-0.20	0.65	0.60	0.71	10.73
	Return 1 year										
1 star	90.88	26.37	15.29	00.78	-42.67	38.41	90.87	-18.50	16.97	18.14	30.28
2 star	11.67	27.78	17.84	30.64	-44.82	35.83	12.17	-15.02	18.70	18.25	20.65
3 star	12.58	28.26	21.40	20.60	-45.07	34.67	12.42	-14.62	19.09	19.56	40.15
4 star	14.58	30.01	22.58	1.27	-45.84	34.69	13.06	-14.42	20.00	19.22	40.38
5 star1	17.04	30.59	24.94	0.56	-47.02	31.20	13.63	-12.06	20.50	19.37	50.76

Table 4- Conditional 1 year performance by Rating and Year

This table reports the values of the dependent variables considered in the analysis obtained from Morningstar direct database. Sharpe Ratio is calculated in an annual basis, Alpha is the beta-adjusted return over one-year period, Return is the net yearly return.

4.2. Models

The first method we use to examine the out-of-sample predictive performance is panel data regression model (Pooled and Random effects). The methodology based on panel data can control individual effects with advantages like the reduction of collinearity and efficiency, among others (Baltagi, 2010). In this sense, we estimate the following equations:

$$Y_{it} = \alpha_i + \beta_1 D4_{it} + \beta_2 D3_{it} + \beta_3 D2_{it} + \beta_4 D1_{it} + \sum_i Year_i + \sum_j Category_j + \varepsilon_{it}$$

where:

Y_i= performance metric for fund i.

i= 1 through N, where N is the total number of funds in the sample.

$D4$ = 1 if the fund received a 4-star overall Morningstar rating, 0 if not.

$D3$ = 1 if the fund received a 3-star overall Morningstar rating, 0 if not.

$D2$ = 1 if the fund received a 2-star overall Morningstar rating, 0 if not.

*D*1 = 1 if the fund received 1-star overall Morningstar in the year, 0 if not.

Year = dummy time variables

Category= dummy Morningstar categories.

α_i and β_1,β_2,β_3,and β_4 are parameters of the regression and ε_i the term error.

We also use this methodology with 1 and 3 year lags, in order to evaluate the predictive ability of Morningstar ratings for medium and short-term performance. For each term, we use specific measures and data to avoid autocorrelation problems. First, we estimate the models using exclusively ratings to assess the effect of using only this variable for funds selection. Later, we include other control variables, like expenses, size or manager experience.

Instead of Panel data and based on Chen and Huang (2011), we also use the quantile regression to extend the regression model to conditional quantiles of the different performance metrics because it is more appropriate for a heterogeneous mutual fund universe where strategies and objectives can vary. This model let us capture information about the coefficients at different quantiles of the dependent variable given the set of endogenous variables (star rating). In addition, the conditional quantile regression developed by Koenker and Bassett (1978) deals well with skewed distributions of fund performance. In particular, we adopt the bootstrapping method proposed by Efron (1979) and implemented in the software Stata 12.

Given Y_i as the different performance metrics used in this paper, and X_i as a vector of exogenous variables representing the rating of the fund, the quantile model can be written as:

$$y_i = X_i^{'}\beta_\phi + u_{\phi i}$$

Assuming that:

$$Quant_\phi(y_i|X_i) = X_i^{'}\beta_\phi$$

$$Quant_\phi(u_{\phi i}|X_i)=0$$

4.3. Results for rating models

The estimation of panel data models presented in Table 5 shows the results of regressing the risk adjusted return with the analyst indicators variables after twelve months since the rating is available, through a robust random effects panel data model. As in previous works, the models were initially estimated without control variables to evaluate the effect of selecting funds based exclusively on the rating analyst. Subsequently, other explanatory variables such as costs, size and management experience have been included. In all cases, category and time control variables have been included. As you can see, the results depend on the metric used, but in general, only funds classified as gold show better 12-month performance than the "Not recommended" funds in terms of Sharpe's ratio. Over 36 months the results are more disappointing and the only significant sign is negative. Therefore, our analysis reveals the inability of analyst ratings to identify funds that outperform their peers except in the case of gold funds, where the results show a better performance than those classified as not recommended. The differences with previous studies may be due to the fact of considering in our study a different sample focused exclusively on US, the panel data methodology and the different period considered.

Table 5 shows that most of rating lagged variables are very significant over the different performance measures within one year. In general, funds that have a better rating one year ago, on average over-performed the subsequent lower grade in the next period. The level of fit observed in most of the estimated models indicates that the rating in addition to have predictive power explains an important part of the fund´s performance. In addition, this predictive power is also noted for yearly total return, indicating that those funds with better risk-return also get superior absolute returns. Our results support, in general terms, the performance persistence of quantitative ratings in the short-term (1 year) indicating its validity as a criterion for selecting funds. On average 1-star ratings obtain -1.02% Alpha than those rated as 5-stars. This difference is also true for the rest of the ratings with values that range from -0.24% for 4 stars to 0.87% for 2 stars. The above is also true for the Sharpe ratio where the profitability of higher-ratings funds outperforms those of lower ones throughout the period analysed. Again, the predictive ability is maintained in the case of Annual Return. However, the differences are not as clear when they are compared with the best funds (4stars), resulting only significant for the Sharpe ratio. In this regard, funds with better ratings (4

or 5 stars) seem to have a better ex-post behaviour, being especially noticeable for those with 1 and 2 stars. Our findings are in line with previous research made by Ferson and Schadt (1996), Morey and Gottesman (2006), Antypas *et al.* (2009), Müller and Weber (2014) and Meinhardt (2014); where they obtain some evidence in behalf of performance persistence.

Table 5- Panel Data with 1 lag (rating overall)

	Pooled			Random		
Variable	**Alpha**	**Sharpe**	**Return**	**Alpha**	**Sharpe**	**Return**
l1.stars4	-0.3055	-0.1396***	-0.1975	-0.2425	-0.1295***	-0.2775
L1.stars3	-0.8635***	-0.2531***	-0.6172***	-0.7579**	-0.2383***	-0.7796***
L1.stars2	-1.0345***	-0.3458***	-0.7152***	-0.8799***	-0.3250***	-0.9747***
L1.stars1	-1.2201***	-0.4799***	-1.6881***	-1.0219*	-0.4493***	-2.0720***
yr2004c	1.1685***	-1.7454***	6.2224***	-1.1630***	-1.7417***	6.2118***
yr2005c	0.7394*	-0.1804***	21.5354***	0.7406***	-0.1775***	21.5385***
yr2006c	-1.2699***	0.2447***	14.8326***	-1.2653***	0.2486***	14.8226***
yr2007c	1.6178***	-0.2375***	-1.8108***	1.6255***	-0.2329***	-1.8267***
yr2008c	5.2850***	-2.2675***	- 47.3888***	5.2937***	-2.2637***	-47.4052***
yr2009c	3.1783***	-1.9710***	26.5907***	3.1712***	-1.9676***	26.5781***
yr2010c	3.6807***	-1.7810***	6.7910***	3.6746***	-1.7783***	6.7846***
yr2011c	-0.2465	-1.0265***	-17.8362***	-0.2471	-1.0243***	-17.8429***
yr2012c	4.0070***	-1.1377***	14.9007***	4.0044***	-1.1367***	14.9021***
yr2013c	3.9122***	-0.8914***	17.4709***	3.9048***	-0.8912***	17.4746***
Largecapblend	-1.2183***	-0.0662***	-1.4546***	-1.2444***	-0.0658***	-1.4521***
Largecapgrwth	0.7505	0.0749***	0.1501	0.7384	0.0795**	0.1458
Largecapvalue	-1.6101***	-0.1492***	-2.2181***	-1.6335***	-0.1536***	-2.2043***
Midcap	-0.0815	0.1446***	3.0937***	-0.0962	0.1428***	3.1154***
Smallcap	0.3016	0.1772***	4.4938***	0.29	0.1767***	4.5258***
LargecapexUK	-0.0228	0.0118	0.5125	-0.0443	0.0125	0.507
SmallmidexUK	1.8133*	0.1417***	3.1985***	1.8002*	0.1483***	3.2359**
Eurozoneflexcap	-2.6167***	-0.1790***	-2.2755***	-2.7362	-0.1818***	-2.2626***
Eurozonelargecap	-1.1144***	-0.2035***	-2.7671***	-1.1366**	-0.2088***	-2.7594***
Eurozonemidcap	2.0759*	-0.0073	1.6261**	2.1472	-0.0081	1.6309**
Eurozonesmallcap	1.7764*	-0.0285	0.8476	1.7516	-0.03	0.8754
cons	-0.4619	1.7133***	5.6667***	-0.5264	1.7005***	5.8201***
N	4435	8776	8860	4435	8776	8860
r2_w	-	-	-	0.1771	0.9317	0.9184
Rho	-	-	-	0.0867	0.2384	0.00
Chi-square (p)	0.000	0.000	0.000	-	-	-

This table reports the coefficients for Panel Data pooled and random models for different performance measures. Alpha is the beta-adjusted return over a one-year period; Sharpe is the yearly risk-adjusted return and, Return is the total one year net return. L1. Star is the one year lagged variable representing the rating of the mutual fund and yr* are the time dummies variables and finally, Largecapblend, Largecapgrwth, Largecapvalue, Midcap, Smallcap, LargecapexUK, SmallmidexUK, Eurozoneflexcap, Eurozonelargecap, Eurozonemidcap and Eurozonesmallcap are dummies to control for categories. N is the number of observations, r2 the pseudo-squared fit measure, Rho is the fraction of variance due to individual effects and Chi-square (p) is the p-value associated to the Chi-square significance test. *Significant at 10%; ** significant at 5% and * significant at 1%.**

As can be seen in the table below (Table 6), the results in the 3 lags pooled and random effects models confirm previous achievements in the pooled model and partially in the random effects. However, the differences in performance reduce considerably and in many cases, they aren´t significant for the random effects model. The results support the significance in shorter time periods, and to a lesser extent, in longer periods. Something that seems logical since in the long term it is likely that the fund rating will change, and consequently, this can affect their future performance.

Table 6 - Panel Data with 3 lags (rating overall)

	Pooled			Random		
Variable	**Alpha**	**Sharpe**	**Return**	**Alpha**	**Sharpe**	**Return**
L3.stars4	-0.1861	-0.0430**	-0.0186	0.0006	-0.0314	0.1198
L3.stars3	-1.0433***	-0.0881***	-0.7414***	-0.7006*	-0.0620***	-0.4324
L3.stars3	-1.2857***	-0.1000***	-0.7216**	-0.7095	-0.0522**	-0.1951
L3.stars1	-1.7105***	-0.1982***	-1.7070***	-1.0025	-0.1276***	-0.9342**
yr2004c	-0.8695***	0.6627***	7.7480***	0	1.5507***	21.9404***
yr2007c	0.2321	-0.8898***	-14.2031***	1.0776***	0	0
yr2010c	-	-	-	0.8142***	0.8861***	14.1835***
Largecapblend	-0.5737	-0.0094	-0.5852*	-0.6119	-0.0118	-0.587
Largecapgrwth	1.1552*	0.1480***	1.2841**	1.1384*	0.1424***	1.2674*
Largecapvalue	-1.1131**	-0.0752***	-1.3074***	-1.1195**	-0.0832**	-1.3268***
Midcap	-0.0681	0.1084***	2.7028***	-0.0483	0.1156***	2.7358***
Smallcap	0.0362	0.1292***	3.5137***	0.0506	0.1309***	3.4996***
LargecapexUK	0.2941	0.0911***	1.4438***	0.2126	0.0839***	1.4289***
SmallmidexUK	1.8214*	0.1209**	3.1604***	1.6888	0.1247*	3.1690***
Eurozoneflexcap	-0.8713	-0.1648***	-2.2805***	-1.3277	-0.1764***	-2.3472**
Eurozonelargecap	-0.2264	-0.0873***	-1.5137***	-0.229	-0.0967***	-1.5212***
Eurozonemidcap	1.2723	0.0057	1.4497*	1.5739	-0.0026	1.3626*
Eurozonesmallcap	1.7389	-0.0066	0.6263	1.6054	-0.0094	0.616
cons	0.8002*	0.5914***	8.3063***	-0.3053	-0.3167***	-6.1738***
N	1510	2989	3017	1510	2989	3017
r2_w	-	-	-	0.1321	0.9131	0.9077
Rho	0.00	0.00	0.00	-	-	-
Chi-square (p)	-	-	-	0.01	0.00	0.00

This table reports the coefficients for Panel Data pooled and random models for different performance measures. Alpha is the beta-adjusted return over a one-year period; Sharpe is the yearly risk-adjusted return and, Return is the total one year net return. L3.Star is the tree year lagged variable representing the rating of the mutual fund and yr* are the time dummies variables and, finally, Largecapblend, Largecapgrwth, Largecapvalue, Midcap, Smallcap, LargecapexUK, SmallmidexUK, Eurozoneflexcap, Eurozonelargecap, Eurozonemidcap and Eurozonesmallcap are dummies to control for categories. N is the number of observations, r2 the pseudo-squared fit measure, Rho is the fraction of variance due to individual effects and Chi-square (p) is the p-value associated to the Chi-square significance test. *Significant at 10%; ** significant at 5% and * significant at 1%.**

Table 7 shows the results of quantile regression with 1 lag confirming previous achievements for the different percentiles. In particular, the positive difference for better ratings in risk-adjusted performance is confirmed for all the levels considered. However, the differences are more significant in the sample of the most profitable funds and those who are around average, than in the 75 percentile (Q75), when we take the total return.

Table 7- Quantile regression for 1 and 3 lags (rating overall): Equation 1

Variable	Alpha (Q25)	Sharpe (Q25)	Return (Q25)	Alpha (Q50)	Sharpe (Q50)	Return (Q50)	Alpha (Q75)	Sharpe (Q75)	Return (Q75)
l1.stars4	-0.7856**	-0.1031***	0.2001	-0.6863***	-0.1494***	-0.4111	-1.2358***	-0.1994***	-0.7705**
L1.stars3	-1.2799***	-0.1953***	-0.0843	-1.4531***	-0.2599***	-1.1650***	-2.2646***	-0.3041***	-1.6594***
L1.stars3	-1.4429***	-0.2989***	-0.5981	-1.5744***	-0.3502***	-1.4717***	-2.5332***	-0.3958***	-1.9874***
L1.stars1	-2.9872***	-0.4588***	-2.2046***	-2.4037***	-0.4694***	-2.4399***	-1.8628***	-0.4967***	-1.9123***
yr2004c	-2.7910***	-1.5629***	-	-1.6996***	-1.5510***	-	-0.8632**	-1.6013***	-
yr2005c	-	-	15.3620***	-	-	15.3565***	-	-	15.1855***
yr2006c	-3.0676***	0.4041***	8.5395***	-1.8229***	0.4826***	8.6942***	-0.5770**	0.4987***	8.6621***
yr2007c	-0.3445	-0.0779***	-8.3565***	0.4629	-0.0082	-7.3436***	2.1939***	0.0554**	-5.1723***
yr2008c	2.1939***	-2.0514***	-53.7852***	3.7600***	-2.0556***	-53.0733***	6.4822***	-2.0776***	-52.2483***
yr2009c	0.9705***	-1.7414***	17.9727***	2.0200***	-1.7384***	19.7295***	3.8944***	-1.7856***	22.2340***
yr2010c	1.1888***	-1.5382***	-2.8859***	2.3813***	-1.5638***	0.3644	3.7456***	-1.6391***	2.7241***
yr2011c	-1.4034***	-0.8576***	-24.1165***	-0.6318**	-0.8110***	-23.1419***	0.4001*	-0.8188***	-22.1111***
yr2012c	2.6613***	-0.9402***	8.8923***	2.6910***	-0.9146***	9.0832***	3.6628***	-0.9592***	9.3028***
yr2013c	2.3230***	-0.6577***	10.8681***	2.7629***	-0.6576***	11.3616***	4.2652***	-0.7141***	12.1108***
yr2014c	-0.7831***	0.1710***	-5.1934***	-0.4341*	0.2197***	-5.3938***	-0.1749	0.2416***	-5.7037***
Largecapblend	0.8840*	-0.0154	0.8613***	-0.717	-0.0668***	-0.9523***	-2.6096***	-0.1168***	-2.6583***
Largecapgrwth	1.6389***	0.0156	0.6766*	0.7471	0.0603**	0.3161	0.4759	0.0791*	0.0175
Largecapvalue	0.1818	-0.0775***	-0.651	-1.1429**	-0.1275***	-1.4526***	-2.9139***	-0.1699***	-2.7086***
Midcap	0.6077	0.1284***	1.4710**	0.0522	0.1691***	3.9220***	-0.845	0.1449***	4.9620***
Smallcap	0.7962	0.1692***	1.2280**	0.5834	0.1823***	4.7041***	0.0315	0.2024***	7.8527***
LargecapexUK	1.8062***	0.0791***	3.2123***	0.456	0.0027	1.1272***	-1.1739	-0.0666***	-0.7071*
SmallmidexUK	2.8893***	0.1229***	1.6649*	2.0117	0.0968	0.4185	2.3599	0.1651***	5.2740**
Eurozoneflexcap	-1.8575	-0.1502***	-2.1267***	-1.4704	-0.1393***	-1.8200***	0.9037	-0.1591***	-1.5396**
Eurozonelargecap	0.8617	-0.1405***	-0.4805	-0.6594	-0.1855***	-1.8067***	-2.3432***	-0.2132***	-3.5288***
Eurozonemidcap	0.7202	-0.007	0.8945	4.5564***	0.0098	1.1887	2.6893*	0.0236	2.7474**
Eurozonesmallcap	2.1983	0.0015	-0.1535	1.8732*	0.0021	0.4545	0.6663	-0.0251	2.3513**
_cons	-2.6679***	1.2847***	6.9382***	0.464	1.4816***	11.3829***	4.1102***	1.7155***	15.3961***

This table reports the coefficients for Quantile regression. Alpha is the beta-adjusted return over a one-year period; Sharpe is the yearly risk-adjusted return and, Return is the total one year net return. L1.Star is the one year lagged variable representing the rating of the mutual fund and yr* are the time dummies variables and, finally, Largecapblend, Largecapgrwth, Largecapvalue, Midcap, Smallcap, LargecapexUK, SmallmidexUK, Eurozoneflexcap, Eurozonelargecap, Eurozonemidcap and Eurozonesmallcap are dummies to control for categories. *Significant at 10%; ** significant at 5% and * significant at 1%.**

4.4. Results with net expenses and other variables

A possible explanation for our result could be the lower costs associated with the best ratings. Carhart (1997) find that the ratio of costs and the rotation of the portfolio are greater and negatively related to performance. The ratio of costs appears to reduce profitability by more than 1% a year, while the rotation reduces it to around 0.95%. Another important author, Gruber (1996), shows the same negative link between cost and profitability and how in the funds with the best performance, the ratio increases over time more slowly than in those with worse results. More recently Nanigian (2012) indicates why mutual fund expenses matter. The author found that there exists a negative relationship between expenses and performance. Authors found that expenses are lower when funds are offered to a sophisticated clientele of investors. Thus, we estimate the following model, where we include the expenses and other explanatory variable such as age, turnover and size:

$$Performance_{it} = \alpha_i + \beta_1 D5_{it-1} + \beta_2 D4_{it-1} + \beta_3 D3_{it-1} + \beta_4 D2_{it-1} + \sum_{t=1}^{11} \beta_i \, DYEAR_{It} + \beta_{12} EXPENSES_{It} + \beta_{13} AGE_{it} + \beta_{14} TURNOVER + \beta_{15} ASSETS + \varepsilon_{it}$$

The results of the estimated models show how expenses have an inverse relationship with the profitability and represent around 0.9% of annual return (Table 8). However, the ratings are still significant in explaining performance in most cases, and mainly for risk adjusted measures. This result indicates that costs are not the only factor that determines explanatory power of quantitative ratings. Other variables, such as the age of the fund and the turnover, reduce the performance in some cases. However, the size of the fund has a positive effect in general terms. This would indicate the need to consider in the selection of funds other variables such as size, age and management costs.

Table 8- Panel Data with all variables by 1 lag (Overall)

	Pooled			Random		
Variable	**Alpha**	**Sharpe**	**Return**	**Alpha**	**Sharpe**	**Return**
l1.stars4	-1.4131***	-0.1225***	-0.7364	-1.4684***	-0.1149***	-0.6207
L1.stars3	-2.2640***	-0.2012***	-1.1863***	-2.3434***	-0.1912***	-0.9945**
L1.stars2	-1.8683***	-0.2777***	-0.6238	-1.9817***	-0.2659***	-0.3593
L1.stars1	-0.734	-0.3896***	-1.1933*	-0.8277	-0.3769***	-0.7607
NetExpense_	-0.2962	-0.0119**	-0.8931***	-0.2933	-0.0104	-0.9010***
Turnover_	-0.0017*	-0.0001***	-0.0009	-0.0016	-0.0001***	-0.0009
LogNetAss	-0.0533	0.0070***	0.0144	-0.0537	0.0081**	0.02
Yearsfund	0.0098	-0.0013	-0.0174	0.0114	-0.0015	-0.0219
yr2004c	-2.4684	-1.8561***	3.2829	-2.6098	-1.8491***	3.7536*
yr2005c	3.8044**	-0.1441**	25.8385***	3.8297**	-0.1455	25.8388***
yr2006c	-1.1462	0.3255***	17.0549***	-1.161	0.3275***	17.0593***
yr2007c	1.4797*	-0.2056***	-0.8712	1.4772*	-0.2032***	-0.852
yr2008c	5.0947***	-2.3058***	-47.1849***	5.0773***	-2.3040***	-47.1446***
yr2009c	3.4629***	-1.9979***	28.1477***	3.4575***	-1.9967***	28.2026***
yr2010c	3.8065***	-1.8037***	8.0594***	3.8016***	-1.8024***	8.1135***
yr2011c	0.3353	-1.0126***	-17.1657***	0.3278	-1.0110***	-17.1202***
yr2012c	4.2355***	-1.1001***	15.1712***	4.2375***	-1.0986***	15.2097***
yr2013c	4.2183***	-0.8903***	17.6879***	4.2283***	-0.8905***	17.6966***
Largecapblend	-1.5537***	-0.0936***	-2.3848***	-1.5516***	-0.0920***	-2.3702***
Largecapgrwth	0.0676	0.0481	-0.0649	0.0518	0.0475	0.0217
Largecapvalue	-2.5555***	-0.1501***	-3.2913***	-2.5685***	-0.1477***	-3.2451***
Midcap	-0.3856	0.1082***	3.5169***	-0.3901	0.1105***	3.5497***
Smallcap	-0.9794	0.1017***	3.2328***	-0.9862	0.1040**	3.2713***
largecapexUK	-1.7993**	-0.0608*	-2.4931***	-1.8102**	-0.0573*	-2.4898***
smallmidexUK	1.3525	0.0834	2.7394**	1.3537*	0.0856	2.8351
Eurozoneflexcap	3.5669**	-0.0985***	-2.1958**	3.5972*	-0.0994**	-2.1989
Eurozonelargecap	-1.9734***	-0.1805***	-5.1526***	-1.9705***	-0.1798***	-5.1665***
Eurozonemidcap	2.2177	-0.0053	0.9101	2.2297	-0.0032	0.9242
Eurozonesmallcap	5.4465***	0.0783*	3.1451***	5.4474**	0.0835	3.2616
_cons	2.3525*	1.6277***	8.1996***	2.4009	1.5982***	7.9444***
N	1515	2545	2589	1515	2545	2589
r2_w	0.00	0.00	0.00	0.1108	0.937	0.9333
Rho	-	-	-	0	0.3071	0.0154

This table reports the coefficients for Panel Data pooled and random models for different performance measures. Alpha is the beta-adjusted return over a one-year period; Sharpe is the yearly risk-adjusted return and, Return is the total one year net return. L1.Star is the one year lagged variable representing the rating of the mutual fund and yr* are the time dummies variables and, finally, Largecapblend, Largecapgrwth, Largecapvalue, Midcap, Smallcap, LargecapexUK, SmallmidexUK, Eurozoneflexcap, Eurozonelargecap, Eurozonemidcap and Eurozonesmallcap are dummies to control for categories. N is the number of observations, r2 the pseudo-squared fit measure, Rho is the fraction of variance due to individual effects and Chi-square (p) is the p-value associated to the Chi-square significance test. *Significant at 10%; ** significant at 5% and * significant at 1%.**

Table 9- Panel Data with all variables and 3 lags (Overall)

	Pooled			Random		
Variable	**Alpha**	**Sharpe**	**Return**	**Alpha**	**Sharpe**	**Return**
l1.stars4	-1.1886**	-0.0625**	-0.517	-0.8959	-0.0532	-0.0287
L1.stars3	-2.1565***	-0.0890***	-1.1234**	-1.7901***	-0.0781**	-0.4575
L1.stars2	-1.9490***	-0.0807**	-0.7985	-1.5052**	-0.0677*	-0.0386
L1.stars1	-1.0794	-0.0632	-0.3168	-0.7239	-0.0438	0.9906
NetExpense_	-0.0799	-0.0249**	-0.223	-0.1604	-0.0254***	-0.2528
Turnover_	-0.0046***	-0.0002***	-0.0023**	-0.0044***	-0.0002*	-0.002
LogNetAss	0.1019	0.0174***	0.2506***	0.119	0.0177***	0.2762***
Yearsfund	-0.0012	-0.0027**	-0.0612***	-0.0103	-0.0030*	-0.0744***
yr2004c	-0.8724	0.7022***	8.0487***	0	0	14.4718***
yr2007c	-0.3323	-0.9262***	-14.3609***	0.4854	-1.6317***	-8.0098***
yr2010c	-	-	-	0.6833	-0.7073***	6.2775***
Largecapblend	-1.2420**	-0.1047***	-1.4279**	-1.4032**	-0.1068**	-1.5462**
Largecapgrwth	0.2181	0.0482	0.3223	0.1413	0.0461	0.2048
Largecapvalue	-1.8029**	-0.1504***	-2.1754***	-1.9825***	-0.1501***	-2.1537**
midcap	-1.3453	0.0822	2.7726***	-1.5203	0.0835	2.8267**
smallcap	-1.9258***	0.051	2.6142***	-2.0120**	0.0524	2.7017***
largecapexUK	-0.8448	-0.0074	0.4092	-0.9428	-0.008	0.4248
smallmidexUK	4.0911***	0.0975	3.0409**	4.0745***	0.1023	3.1702***
Eurozoneflexcap	4.4160**	-0.0335	-1.4089	4.3844***	-0.0379	-1.5757
Eurozonelargecap	-0.6275	-0.1712***	-2.8972***	-0.7935	-0.1745***	-3.0774***
Eurozonemidcap	3.1912**	0.0483	2.3518*	2.9340**	0.0484	2.3374**
Eurozonesmallcap	7.8447***	0.0188	0.6598	7.6019***	0.0156	0.496
_cons	1.6124	0.5093***	7.0640***	0.6078	1.2052***	0.00
N	384	619	631	384	619	631
r2_w	-	-	-	-	0.9343	0.9195
rho	0	0	0	-	-	-
Chi-square (p)	-	-	-	0.0	0.05	0.1931

This table reports the coefficients for Panel Data pooled and random models for different performance measures. Alpha is the beta-adjusted return over a one-year period; Sharpe is the yearly risk-adjusted return and, Return is the total one year net return. L1.Star is the one year lagged variable representing the rating of the mutual fund and yr* are the time dummies variables and, finally, Largecapblend, Largecapgrwth, Largecapvalue, Midcap, Smallcap, LargecapexUK, SmallmidexUK, Eurozoneflexcap, Eurozonelargecap, Eurozonemidcap and Eurozonesmallcap are dummies to control for categories. N is the number of observations, r2 the pseudo-squared fit measure, Rho is the fraction of variance due to individual effects and Chi-square (p) is the p-value associated to the Chi-square significance test. *Significant at 10%; ** significant at 5% and * significant at 1%.**

4.5. Results with downside risk

In addition to the performance, an aspect appreciated by investors is the ability of funds to deal with periods of extreme conditions. Therefore, the downside risk, measured by the maximum empirical loss suffered by the fund at different confidence levels has been included. As it can be shown in Table 11, in general, the best ratings have a better behaviour in terms of empirical VaR with a confidence level of 99%. This finding is important, because investors who select mutual funds based on ratings, will obtain a better performance but, at the same time, less extreme losses. However, even

though differences are positive for good rated mutual funds, they are very narrow in certain years. Nevertheless, it seems that the investment in good rated funds can help to preserve the investors´ wealth better in normal market conditions but not in the case of big downfalls where the losses have been greater than those lower rated.

When we analyse the downside risk conditioned to the rating of the fund three years ago (Table 10), we note that in general the coefficients are significant for the empirical VaR 99%. These results indicate that the worst category funds have higher down risk in relation to 5 stars category. This means that investment in funds with good rating also protects against extreme falls. It is important to note in relation to the categories, that those who behaved better from the point of view of performance (small and midcap), they are not in terms of extreme losses.

Table 10.- Panel Data with all variables and 3 lags (Overall)

Variables	VaR
L3.stars4	0.4410**
L3.stars3	0.4560***
L3.stars2	0.3759*
L3.stars1	0.8253***
yr2004c	0.005
yr2007c	9.2771***
yr2010c	4.4573***
Largecapblend	-0.6093**
Largecapgrowth	-1.3938***
Largecapvalue	-0.3218
Midcap	0.6032
Smallcap	1.3443***
LargecapexUK	0.1918
SmallmidexUK	1.0327
Eurozoneflexcap	1.8871**
Eurozonelargecap	0.3109
Eurozonemidcap	1.1951***
Eurozonesmallcap	0.2446
_cons	5.3729***
N	2101
r2_w	0.8267
Rho	0.2768

This table reports the coefficients of random models for Value at Risk. VaR is the value at risk three years measure. L3.Star is the three years lagged variable representing the rating of the mutual, yr* are the time dummies variables and, finally, Largecapblend, Largecapgrwth, Largecapvalue, Midcap, Smallcap, LargecapexUK, SmallmidexUK, Eurozoneflexcap, Eurozonelargecap, Eurozonemidcap and Eurozonesmallcap are dummies to control for categories. N is the number of observations, r2 the pseudo-squared fit measure and Rho is the fraction of variance due to individual effects. *Significant at 10%; ** significant at 5% and * significant at 1%**

Table 11- Monthly three years Conditional VaR by Rating and Year

	VaR 99%										
	2004	**2005**	**2006**	**2007**	**2008**	**2009**	**2010**	**2011**	**2012**	**2013**	**2014**
1 star	16.08	70.50	50.66	50.96	15.51	20.14	19.45	12.60	12.95	12.39	70.14
2 star	14.25	60.61	50.22	50.74	15.77	17.32	17.42	11.43	11.13	11.28	60.67
3 star	12.78	60.33	50.82	60.38	15.66	15.95	15.98	11.10	10.36	10.23	60.33
4 star	11.67	5.87	60.06	6.55	16.35	15.06	14.91	10.63	90.79	90.63	60.11
5 star	90.85	50.429	50.73	60.62	17.10	13.26	13.56	90.53	90.00	80.95	50.55

This table reports monthly three years Conditional VaR by Rating and Year- VaR is the empirical monthly value at risk over a three years' period with a confidence level of 99%.

5. Robustness

We conducted some additional robustness tests to check the consistency of our results and to provide other complementary analysis. We made the analysis of the regression by year where results are included in the three first tables of Appendix I (Table 37, 38 and 39). In general, the results obtained confirm the best out of sample performance of the best ratings in a greater number of periods. However, the cross-section models allow us to observe that during the period of the crisis, and in particular in 2008, there are some alterations in the usual pattern. Thus, in the 1-year Alpha, 5-star rating funds performed worse than the three- or two-star rating, and in 2008 or 2009, there are no significant differences compared to other lower rating categories. The same happened with the Annual Return, which in 2007 and 2009 was lower in the 5 star funds. In any case, the results confirm that although there may be some years in which the better ratings have a worse out of sample performance, it is much higher the number of periods in which the result is better for the higher ratings. In this sense, it is still confirmed that the strategy of investing in better rated funds may lead to better future yields. On the other hand, using three years rating (instead of overall), we obtain similar results when we use the one year performance of the different indicators (Table 40).

6. Conclusions

Many investors select their investments in mutual funds based exclusively on the rating. In this sense, we wanted to verify if the exclusive use of this criterion allows making good decisions in terms of performance and downside risk. The data has been collected from Morningstar Direct database covering the period 2003 and 2014. We selected European equity funds and we used Rating Overall and also three years ratings to check the robustness of our estimates. Our conclusions support the ability of quantitative ratings to select funds that will behave better in terms future performance. Our results are in line with some previous empirical evidence found in Morey and Gottesman (2006), Müller and Weber (2014), and Meinhardt (2014). In this way, we have found that on average, funds with a lower rating have a worse performance in terms of risk adjusted measures and Annual Return for the following year. The strongest predictability is observed one year ahead but it is also good for three-years. The inclusion of other variables such as costs, size and age reflects the importance of considering other variables for fund´s selection. Nevertheless, the ratings are still

significant in explaining performance, indicating that costs are not the only factor that determines the predictive power of quantitative ratings. Finally, the best ratings perform better in terms of VaR showing that the investment in good rated funds can help to preserve the investors´ wealth better. Our results support the use of ratings in the investment funds selection process, accompanied by other quantitative variables. On the other hand, the greater significance achieved in the short term advises the review of portfolios on an annual basis. Finally, the inclusion of qualitative factors can help improve the selection process of investment funds.

Chapter III: Does Morningstar Analyst Rating matters for mutual funds?

1. Introduction

Mutual funds grew explosively in recent times and the researches on this topic have multiplied. In fact, companies, banks, investors, and others are caring for mutual funds so it is a matter of interest that goes beyond the academic world. To make life easier for investors, rating agencies such as Morningstar, Lipper Leaders, MSCI, Standard and Poor's, among others, attached notes to mutual funds to help investors in selecting their funds. In fact, there are investors who make their decisions based exclusively on ratings. This is why many funds use their rating as advertising to attract investors. Most popular ratings are Morningstar Star ratings and some empirical studies as Blake and Morey (2000) and Guercio and Tkac (2008) and among others, have shown that the downgrade or upgrade of quantitative ratings have an influence in the flows of mutual funds. In addition, in the previous chapter we have shown the ability of Stars Rating (backwards looking) to explain the out-of-sample performance in the short term and to preserve the wealth of investors. Authors such as Morey and Gotesman (2006), Müller and Weber (2014) and Meinhardt (2014) have also supported the performance persistence of quantitative ratings but only in the short term.

However, the selection of funds based exclusively on historical performance or quantitative ratings, excludes a set of qualitative factors that can explain future performance. This is why, in addition to quantitative ratings it have appeared ratings based on analyst opinions that evaluate features of mutual funds as: Governance, Process, People, Parent, Board Quality, Corporate Culture, Fees, Manager Incentives or Regulatory Issues. Despite there exists several studies about quantitative ratings, qualitative ratings are not as popular and very few research has been done about the ability to select good funds based on qualitative ratings. In the particular case of Morningstar, there are two alternatives: Analyst Ratings and Stewardship Grade. As far as we know, very few researches have been conducted on this subject. Wellman and Zhou (2008), Ng (2009), Lai, Tiwari and Zhang (2010), Chen and Huang (2011), Gottesman and Morey (2012) and Cao, Ghosh, Goh and Ng (2012*) st*udied the effects on performance and flows of Stewardship Grades. On the other hand, Kamal (2013) and

Armstrong, Genc and Verbeek (2016) are the unique authors that focus their research on Analyst ratings.

Analyst rating was launched in September 2011 by Morningstar and it is a forward-looking measure based on analyst's expectation about the future performance of the mutual fund relative to the peers and for the long term. The rating reflects the valuation of analyst in five dimensions which includes factor like the cost, past performance, quality of management, interest alignment, etc. Armstrong *et al.* (2016) find "higher abnormal flows to funds receiving higher ratings suggesting that the average retail investor values the analyst's subjective views when allocating their wealth". Thus, investors take into account both stars (backward-looking) and analyst (forward-looking) to take their decisions of investment.

In this paper we assess to what extent selecting mutual funds based on ratings criteria has an impact on the financial and risk performance of investors. In particular, we endeavour to answer three essential questions: (1) Do good analyst (forward-looking) ratings outperform non-recommended ones in the short and long term? (2) Do good stars (backward-looking) ratings outperform bad ones in the short and long term? (3) Is useful to combine both ratings in the screening process to identify good future performers?

Previous research on analyst ratings is scarce and, in general, includes all the universe of mutual funds with rating; despite the heterogeneity in terms of investment area, exchange rate risk, period of analysis, etc. This is why our research only contains the analyst ratings for rated funds from August 2012 to August 2016; because previous studies have been done with very limited samples or assuming questionable hypothesis about rating persistence. Further, the results of the literature about the ability of "good analyst rated" to outperform "bad analyst rated" have shown mixed evidence. The empirical research into these questions is of particular interest to asset managers, financial advisors and investors using ratings to take their portfolios decisions.

Our results are in line with Kamal (2013) and Armstrong *et al.* (2016). In this way, we have found a small evidence that, on average, funds with a better Analyst Rating (Gold) have a better performance in terms of risk adjusted measures (alpha and Sharpe). The predictability is observed in several analyses done in one year ahead but not for three-years. This evidence is more relevant in the case of the analysis made by investment style´s category. In the analysis of the pillars in which the analyst ratings are

broken down we do not find evidence that future performance is related to any of these specific dimensions.

The paper is organized as follows. First, we describe Morningstar Analyst and Stewardship Grade methodology. Then, we review and summarise the main existing research about qualitative ratings and mutual fund's performance. In the fourth section, we present the empirical analysis, the statistical models and the main results. Finally, we summarize the main conclusions.

2. Morningstar qualitative ratings.

Morningstar has two systems to classify mutual funds based on qualitative aspects: Morningstar Analyst Rating and Morningstar Stewardship Grade. Stewardship Grade is determined using some quantitative measures, but it is primarily based on qualitative information across five areas: Corporate Culture, Fund Manager Incentives, Fees, Fund Board Quality and Regulatory History. The Morningstar Stewardship Grade for Funds assigns a letter grade from A (best) to F (worst) for each fund. Morningstar Analyst Ratings are forward-looking qualitative and quantitative analyses of mutual fund about five pillars: Process, Performance, People, Parent and Price. Analyst Ratings are based on the convictions that funds will outperform their benchmarks over the long term.

2.1. Stewardship Grade

In 2003, a series of scandals related to the management (late trading, market timing and other irregularities) affected US mutual funds (see for example Bogle, 2010 for a summary). The importance of this issue took a series of regulatory reforms, but also rating agencies began to focus more on the issue. In August 2004, Morningstar launched Fiduciary Grades for funds, in 2005 renamed Stewardship Grades. These grades provide a standard of corporate governance ranging from A (best) to F (worst). Stewardship Grades are calculated as the aggregate scores of five components – Corporate Culture, Board Quality, Manager Incentives, Fees and Regulatory History-. Morningstar (2010) reveal the details of the methodology for the Mutual Fund Stewardship Grade. One important changes in the methodology is that in 2011, Morningstar began to use Stewardship Grades in Analysts Ratings because they have better information about Parent pillar. Another important change is that Stewardship Grade changed the weight of the components corporate culture and fund manager

incentive because they found more predictive power in an empirical study. The new grade now applies to fund companies rather than to individual funds only.

The scoring of the Stewardship Grades makes a scale that`s are graded from A (best funds) to F (Worst fund´s). The grades depending all the time of their specified relative cultures. Funds that have a superior culture will standard; but that haven´t all the best practices receive B. Fund companies that meet industry standards receive D or F grades. The maximum global score is 10 points, and is based on the sum of the five component scores: A (9–10 points). B (7–8.5 points), C (5–6.5 points), D (3–4.5 points) and F (2.5 points or fewer). The Corporate Culture component's maximum score is 4 points. For the Board Quality, Manager Incentives, and Fees components, the maximum score is 2 points, and points are awarded in increments as 0.5 points. For the Regulatory History component, the maximum score is 0 points, and the lowest possible score is -2 points. Regulatory History scores may be reduced in increments as 0.5 points. The components that Morningstar evaluate are the follow:

1) Regulatory Issues: examine any regulatory issues fund with time horizon at the last past 3 years.

2) Board Quality: focus on the quality of fund`s board, looking at multiplies factors.

3) Manager Incentives: evaluate two aspects. The first is the fund ownership, when they pretend understand if the manager has significant investment in the fund, if the funds run by the manager are inappropriate for such a large investment, among others. Second is the compensation structure, if exist incentive programs that encourage a focus a short-term performance or asset growth are viewed less favourably.

4) Fees: this component pretends to understand two aspects. First, If the fund`s expense ratio is below the average for its type of share class and second if the fund`s expense ratio declined meaningfully as assets have grown.

5) Corporate Culture; this component looks at how seriously a firm takes its fiduciary duty to its fund´s shareholders.

2.2. Morningstar Analyst Rating

Morningstar Analyst Rating, created by Morningstar in November 2011, are forward-looking qualitative and quantitative analyses of mutual funds. The Analyst Rating is expressed as metals: Gold, Silver, Bronze, Neutral and Negative. The medals Gold, Silver, Bronze are the notes for better funds (Recommended) and the Neutral and Negative are notes by worst funds (Not Recommended). Analysts evaluate funds based on five key pillars: Process, Performance, People, Parent and Price. These keys pillars are the components that Morningstar analysts believe that may be the predictors of outperform over long-term on a risk-adjusted basis. Thus, Morningstar Medallists—Gold, Silver, or Bronze rating—are funds that Morningstar analysts believe will perform better over time compared to similar investments (category group) in the long run (five years). Morningstar (2011) summaries the methodology:

1) Process: analysts try to understand the strategy and how management has competitive advantage to run the process well and consistently.

2) Performance: analysts try to understand what is fund`s strategy and the pattern logical given its process. Another thing is understanding if a fund has strong risk-adjusted returns over a relevant time.

3) People; understand manager´s talent, tenure and resources.

4) Parent; understand if firm prevail Salesmanship or Stewardship (e.g. create alignment of interests with their clients, good governance or have a long-term investment horizon, charge reasonable fees, etc.).

5) Price; due that Morningstar knows that expenses are one of the better predictors of future outperformance this pillar try to resume if a fund have reduced cost compared with similar funds sold through similar channels.

Each pillar is rated positive, neutral or negative. Morningstar Analyst Ratings are based on overall analysis and ratings of the five pillars. Morningstar (2011) describes Analyst Ratings in the follow order:

1) Gold; fund distinguished between the five pillars and have guarantee of highest level of conviction that will perform better over time compared to similar investments.

2) Silver, fund with sufficient level of conviction to guarantee a positive rating and the advantages are bigger than disadvantages in the five pillars.

3) Bronze; fund witch notable advantages across several, but not all of pillars.

4) Neutral; funds that don't have a strong positive or negative conviction. but they aren't likely to seriously underperform their relevant performance.

5) Negative; fund that Analyst Rating consider an inferior offering to its peers and that has reason likely to significantly underperform (e.g. high fees or an unstable management team).

Morningstar may also use two other designations in place of a rating: Under Review and Not Rateable. Under Review means that the fund requires further review to determine the impact on the rating. Not Rateable is used in the case there are no relevant comparators.

Haslem (2014) summaries several works published in the Morningstar Fund Investor, a monthly newsletter of Morningstar, that explain several aspects of the construction and implications of Analyst Ratings and Stewardship Grades (among others). At the same time, the author summarizes the main research assessments of Morningstar ratings on mutual fund performance.

3. Previous research.

There is a debate in literature about the power of ratings to predict future performance. Most of the studies employ quantitative ratings in order to look at the capacity of the ratings by choosing the best funds. The quantitative aspects of ratings have some limitations. Quantitative ratings cannot quantify aspects of qualitative ratings. It is true that when you build ratings based on historical performance in some way implied the ability of managers, as well as other qualitative aspects. However, there is no evidence that past performance has this ability to capture the ability of managers to get results above the market. On the other hand, there are aspects that have to do with ethics, correct procedures, appropriate strategies, as well as whether fund managers are hedging the fund's wealth in the medium and long term. Certain qualitative aspects such as quality of management and procedures cannot be captured by past performance and it is in this sense that qualitative ratings can give very useful information to choose the best funds.

Tufano and Sevick (1997) is the first empirical study examining quality of the board of directors at mutual funds. They examine the relationship between the composition and compensation of boards of directors of U.S. mutual funds and the fees charged to their investors. In the last years, authors as, Del Guercio, Dann and Partch (2003), Qian (2006), Meschke (2007), Khorana, Servaes and Wedge (2007), Khorana,

Tufano and Wedge (2007), Ferris and Xuemin (2007), Boyd and Yilmaz (2007), Trahan (2008), Evans (2008), Kong and Tang (2008), Cremers, Driessen, Maenhout and Weinbaum (2009), Adams, Mansi and Nishikawa (2010), Chou, Ng and Wang (2011), Ding and Wermers (2012), Hazenberg (2012), Kryzanowski and Mohebshahedin (2016) or Mamatzakis and Xu (2017) investigated the quality of governance at mutual funds and they found that in general, funds with better governance obtain better future performance.

There are other papers that focus on the influence of Stewardship Grade or Analyst Rating on fund flows. Wellman and Zhou (2008) found that investors sell funds with poor Stewardship Grades and buy those with good grades. Lai, Tiwari and Zhang (2010) suggest that investors react more strongly to poor fund performance by withdrawing funds when the board quality component of Morningstar's Stewardship Grade is perceived to be bad. Armstrong *et al.* (2016) concluded that higher Analyst ratings (Gold and Silver) receive higher abnormal flows[1], particularly in retail funds. These studies report evidence on the importance for qualitative ratings to investors.

The relative novelty of qualitative ratings of Morningstar causes that, unlike investigation of Star Ratings, there is not too much research on this topic in the literature and just few authors have studied the subject. In the case of Stewardship, we highlight Gerrans (2006), Wellman and Zhou (2008), Ng (2009), Lai, Tiwari and Zhang (2010), Chen and Huang (2011), Gottesman and Morey (2012), Cao *et al.* (2012). For Analyst, only Kamal (2013) and Armstrong *et al.* (2016) had studied their effect on future performance.

Gerrans (2006) investigated the relationship between Morningstar Star ratings, a qualitative rating (QL) and their product in the performance of Australian managed funds. QL rating was an assessment of fund administration, investment management, product and company capabilities and strengths reported as Business and Management Strength rating and Sector Strength rating. He employed data form two of the largest fund subcategories Australian Equity Trusts–General (AET) and Superannuation–Australian Equity Trusts (SAET). Data were collected from the Morningstar Total Access CD between August 1996 and February 2001. The results do not provide evidence to support a positive relationship between ratings and four commonly used

[1] Abnormal flows are computed as the difference between the flow to the rated fund and the flow to a "matched" unrated fund within the same style classification using propensity score matching (PSM).

performance measures (geometric monthly return, one- factor alpha, four-factor alpha and Sharpe ratio).

Wellman and Zhou (2008) is the first work in study future performance employing Stewardships. They concluded that there are significant differences in performance between mutual funds that have good Stewardship Grades and those who have bad. Funds with top Stewardship Grade ("A" or "B") outperform those with poor grades ("D" or "F") by 19 to 23 basis points per month over the period analysed January 2001 - July 2004, and by 10 to 16 basis points over the period September 2004 – December 2004. In the 27 months after the announcement of the grades, good funds outperformed bad funds by a significant 10 basis points. They find that among the five stewardship components, only Fees and Board Quality exhibit significant explanatory power. Regulatory History, Manager Incentives, and Corporate Culture show no explanatory power in explaining ex-post risk-adjusted returns.

Ng (2009) examines applying least-squared regressions and multinomial ordered logit regressions the extent to which Morningstar Star Ratings and Morningstar Stewardship Grades can predict future fund performance. In particular, he investigates the combined predictive power of the two ratings in a twelve-month sample (January 2005 - December 2005). His results show that none of the ratings alone possesses strong predictive power, but the combined rating is superior in forecasting future returns.

Lai, Tiwari and Zhang (2010) focus on the board quality component of Morningstar's Stewardship Grade. They find a significant relationship between board quality and performance persistence. For funds with board quality, negative past performance predicts future negative performance and also there is evidence of short-term persistence in positive performance (positive past performance predicts future positive performance).

Chen and Huang (2011) study the relationship between the performance and Stewardship grades of Morningstar using the methodology of the OLS and quantile regression from the 2006 to 2009 (2^{nd} quarter). In the study, the authors used the overall Stewardship Grade and by another hand, two components grades (Manager Incentives and Board Quality). The statistical results indicate Stewardship Grades are strongly positively related to fund performance measured by the Sharpe Ratio. The OLS regression reveals a strong association between overall Stewardship Grade and the fund performance, but it is not so clear that all the components can have the same predictive power. Quantile regressions show that there is a strong relationship with the right tail of

the performance distribution. Authors also show that Stewardship Grades are strongly negatively related to portfolio turnover. Finally, this study shows that there is little ability to predict future performance of Alfas.

Gottesman and Morey (2012) have studied if Stewardship Grade can predict future performances. They tested the capacity of their components to predict risk-adjusted performance of domestic equity funds over the period 2005-2010. The authors used methods that are robust to survivorship bias and they find that corporate culture have little predictive power on future performances. They also find that no one component of Stewardship Grade can predict fund performance consistently.

Cao *et al*. (2012) test the role of Morningstar Stewardship Grade in mutual fund performance. They use data form Morningstar Direct and Centre for Research in Securities Prices (CRSP) Survivorship Bias Free Mutual Fund database over the sample period November 2004 – May 2011. Their findings suggest that corporate governance grades of mutual funds carry information for predicting long-term mutual fund performance.

Research on Morningstar Analyst Rating is recent and limited. To our knowledge only Kamal (2013) and Armstrong *et al.* (2016) have analysed the effect on Morningstar Analyst on future performance.

Kamal (2013) studied Morningstar Analysis Ratings at July 2013 (1,159 individual mutual funds: equity fund, fixed-income funds, etc.) and concluded that there is a significant positive relationship between these ratings and the future performance as measured by the 3-year Alpha applying OLS and quantile regression. Results for quantile regression show that for better performing funds; higher Analyst Rating does not necessarily predict better performance in the future. He also found that the People pillar of these ratings has a significant predictive power for funds' future performance. However, the author cautions in her work that "whether the Analyst Ratings can predict future fund performance, we need more data, which is not available as of yet, because these are relatively newer ratings, with a long-term focus. Finally, she found that Sharpe Ratio and Analyst Ratings are significantly positively related to contemporaneous fund performance. The People and Process Ratings are also individually significantly related to the Sharpe Ratio.

Armstrong *et al.* (2016) tried to understand to what extent if Morningstar Analysis Ratings have the power to influence investors in terms of flows, as well as these ratings have the ability to provide above average performances. The authors

employed a sample of 412 equity funds from September 2011 to June 2014 and examined if the analyst ratings contain information about the rated funds' future performance. They computed out-of-sample performance using cumulative risk-adjusted returns measured over the 6, 12, 18, and 24 months following each fund's initial rating and identified out-perform peer funds horizons of up of 18 months or more comparing Gold, Silver and Bronze rated funds with Not Recommended funds. Finally, they showed that a portfolio of Gold rated funds has significantly higher alphas than a portfolio that contains all funds that are not rated Gold. Table 12 summaries the preceding works.

Table 12 - Literature Review of predictive power of Stewardship grades and Morningstar Analyst on future mutual fund performance.

Authors	Sample	Number	Predictive Power	Conclusions
Wellman and Zhou (2008)	*Morningstar Stewardship Grades and CRSP Survivor-Bias Free U.S. Mutual Fund Database Ex-ante data: January 2001 to July 2004. Ex- post date: September 2004 to December 2006*	*367 U.S. domestic equity funds*	*Yes*	*- In the ex-post returns analysis authors find difference between the performance of good and bad funds highly significant. - Funds that are poorly governed, whether due to high fees, regulatory issues, poor boards, poor managerial contracts or poor corporate culture, underperform those that have good governance. -The results are strongly driven by board quality and fees. The effects of regulatory issues, managerial incentives and corporate culture are not significant.*
Ng (2009)	*Morningstar Stewardship Grades and CRSP Survivor-Bias Free U.S. Mutual Fund Database January 2005 – December 2005*	*1,589 funds (494 U.S. domestic stock funds, 409 international stock funds and 686 bond funds)*	*No*	*- For the out-of-sample performance analysis (12 evaluation months form December 2004 to November 2005) funds with poor ratings generally continue to exhibit poor future performance and underperform their peers with better ratings, but differences are not significant. - Better results are obtained if Morningstar Stewardship Grades are combined with Stars Rating, so combined rating is superior to single rating in forecasting future returns.*
Lai, Tiwari and Zhang (2010)	*Morningstar Board Quality Rating CRSP Survivor-Bias Free U.S. Mutual Fund database January 2001 - December 2007*	*461 U.S. domestic equity funds*	*Yes*	*- Board quality plays a critical role in determining whether fund performance persists - For funds with bad boards, negative past performance predicts future negative performance. - For funds with good boards there is evidence of short-term persistence in positive performance.*
Chen and Huang (2011)	*Morningstar Stewardship Grade 2006-2009 (2nd quarter)*	*4,164 U.S. funds*	*No*	*- Their study analyses the effect of Stewardship Grade, and two of these components (manager incentives and board quality) in future performance thought OLS and quantile regressions. - Stewardship Grade have no power both OLS and quantile regressions to predict the 3 -years Alpha - Manager incentives variable is not statistically significant, while board quality turns is positive and significant-to predict future mutual funds' performance.*
Gottesman and Morey (2012)	*Morningstar Stewardship Grade 2005-2009*	*U.S. domestic equity funds It depends on the year 376 (2005) - 365 (2019)*	*No*	*- This paper analyses using two methods to avoid survivorship bias the predicting power of fund performance with corporate culture over a 12-, 24- and 60-month out-of-sample period. - Funds with excellent corporate culture ratings have lower expense and turnover ratios than other funds. Also, funds with better corporate culture ratings generally have substantially longer managerial tenure than other funds. - No evidence that any of the Stewardship components can consistently predict future performance.*
Cao et al. (2012)	*Morningstar Stewardship Grade CRSP Survivor-Bias Free U.S. Mutual Fund database November 2004 -May 2011*	*It depends on the year 826 (2005) - 865 (2011)*	*Yes*	*- This paper examines the predictive power of Morningstar's Star Rating and Stewardship Grade. - Star Rating is associated with good one-year post-rating risk-adjusted return. - Morningstar Star Stewardship score (using, the proposed First Principal Component-FPC- score) provides a powerful tool for fund selection.*
Kamal (2013)	*Morningstar Analyst Ratings 2010 - 2012*	*1,159 mutual funds*	*Yes*	*- Author employs OLS and quantile regression models where Alpha is the dependent variable and the Analyst Rating is the independent variable. - OLS regression shows that higher Analyst Ratings do predict better future performance. - Quantile regression shows that Analyst Rating is significantly positively related to future performance along the entire distribution of the Alpha. However, for better performing funds; higher Analyst Rating does not necessarily predict better performance in the future.*
Armstrong, Genc and Verbeek (2016)	*Morningstar Analyst Ratings CRSP Survivor-Bias Free U.S. Mutual Fund Database September 2011 - December 2012*	*412 equity funds*	*Yes*	*- Authors compute out-of-sample performance using cumulative risk-adjusted returns measured over the 6, 12, 18, and 24 months following each fund's initial rating. - At the 24-month (18-month) horizon, Gold rated funds outperform Not Recommended funds by 4.9% (3.3%) when three-factor risk-adjusted returns are used, and by 4.5% (3.2%) when four-factor risk adjusted returns are used. Comparable results are obtained including control variables. - Silver and Bronze rated funds also outperform Not Recommended funds. - A portfolio of Gold rated funds has significantly higher alphas than a portfolio that contains all funds that are not rated Gold*

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

4. Empirical study

4.1. Data and sample

Morningstar Direct reports from 2011 the analyst rating but the sample is very limited till half 2012 when there are around 200 mutual funds rated for the United States investment area. Previous work has been done with several limitations like: all the universe of mutual funds with rating is included, the heterogeneity in terms of investment area is not considered, mutual funds have different level of exchange rate risk, limited period of analysis, etc. This is why in our research we only include the analyst ratings for rated funds from August 2012 to August 2016 because previous research has been done with very limited samples or assuming questionable hypothesis. We restrict our selection to the funds focused on the investment area of United States, with stars ratings and included in the nine common categories resulting from combining size and value[2]. In line with Armstrong *et al.* (2016) we also eliminate from the selection multiple share class, including only one equivalent for each mutual fund and we eliminate the funds less than two years old and less than five million of net assets, to avoid incubation bias.

The sample contains 10.772 monthly observations and an average of 220 mutual funds rated, with a good representation in each level except for the negative case, where surprisingly, any mutual fund has achieved this grade. Similarly than Armstrong *et al.* (2016) we distinguish between "Not recommended" and "Recommended" mutual funds, where recommend is composed by the categories from Gold to Bronze, and the others are classified into "Not recommended". More than 80% are classified as "Recommended" because they have a gold, bronze or silver rating and this is explained by Morningstar by the fact that they prioritize high quality funds and in general they are bigger and with lower expenses and turnover ratios (see Table 13).

Table 13.- Analyst ratings distribution for the period (August 2012-August 2016)

Rating analyst	Freq.	Percent	Cum.
Gold	1,923	17.85	54.71
Bronze	3,97	36.85	36.85
Silver	2,811	26.1	99.61
Neutral	2,026	18.81	73.51
Under Review	42	0.39	100
Total	10,772	100	

In Table 14 we compare the distribution of analyst ratings and star ratings and we observe that better analyst has a greater proportion of good backward looking mutual funds (stars).

[2] Large value, large blend, large growth, medium value, medium blend, medium growth, small value, small blend and small growth.

Nevertheless, a 30% of four and five stars ratings are classified as neutral and thus, as "Not recommended" by Morningstar analysts. At the same time, 27% of bad historical performance is classified into "Recommended" funds, showing that analysts can expect a good performance in the future despite the bad behaviour in the past.

Table 14. Analyst and star ratings distribution

	Rating Overall				
Analyst	**1**	**2**	**3**	**4**	**5**
Gold	0.6%	3.6%	20.6%	55.1%	20.2%
Silver	0.9%	7.9%	27.7%	39.8%	23.7%
Bronze	1.0%	13.2%	34.4%	38.7%	12.7%
Neutral	4.1%	24.8%	41.2%	22.9%	7.1%
Under Review	3.1%	6.3%	53.1%	25.0%	12.5%

Table 15, summarizes the fund characteristics across the rating categories, sorting the rated funds into the five analyst rating categories from the highest rating of Gold to the lowest rating. We observe that, on average, Gold, Silver and Bronze rated funds are larger than Neutral and Under-review rated funds. Gold mutual funds have the lower levels of expense ratios, low rotation and higher levels of risk adjusted performance than the others mutual funds. At the same time, Neutral mutual funds are the worst in performance and downside risk, measured through Value at Risk, despite lower expenses ratios and bigger size than Silver and Bronze.

Table 15- Fund characteristics by analyst rating

Variable	Gold	Silver	Bronze	Neutral	Under review
ManagerTen_	93.78	97.22	84.95	60.54	36.40
Netexp_	0.70	0.93	0.99	0.91	0.92
Turnover_	20.83	33.08	45.28	62.83	53.84
VaR99_	5.19	5.10	4.96	5.30	4.00
Sharpe12_	1.13	1.10	1.07	0.95	1.65
Sharpe36_	1.09	1.07	1.06	1.00	1.13
Alpha12_	0.13	-0.40	-1.16	-1.77	-0.40
Alpha36_	0.68	0.33	-0.78	-1.94	-0.59
Return12_	18.27	18.47	18.50	17.10	18.23
Return36_	18.48	18.15	18.04	17.56	18.22
LogSize	10,400.00	2,450.00	2,510.00	3,020.00	1,370.00

Note: ManagerTen is the number of months that a manager(s) has been at the helm of the fund. Netexp is the net expense ratio declared. Turnover is the turnover ratio, i.e., the percentage of the mutual fund that have been replaced with other holdings in a given year. VaR99 is the value at risk at a confidence level of 99%. Sharpe12 (sharpe36) is the ratio of Sharpe (the average return earned in excess of the risk-free rate per unit of volatility) calculated over a period of 12 (36) months. Alpha12 (Alpha36) is the excess return of a fund relative to the return of the benchmark index calculated over a period of 12 (36) months. Return12 (Return36) is the net return of a fund calculated over a period of 12 (36) months. LogSize is the logarithm of the assets of the fund.

To measure out-of-sample performance we use two risk-adjusted metrics: Alpha and Sharpe ratio. Table 16 contains the out-of-sample performance and downside risk measures of the mutual funds three years after the initiation. Thus, we try to analyse if those mutual fund

good rated have in general a better ex-post good out of sample performance. We also differentiate between Analyst and Stars ratings to assess the importance of backward looking or forward looking explaining future performance. In the particular case of star ratings, we observe a monotonic decrease across the categories for all the performance metrics, showing that in general, better rated funds have obtained better performance and less downside risk 36 months later. Surprisingly, for the Analyst rated funds we observe that Neutral mutual funds obtain in general greater performance than "Recommended", except in terms of alpha, but with a higher downside risk. Comparing Gold, Silver and Bronze; we observe greater performance for the best ratings and the contrary in terms of downside risk, but with minor differences.

Table 16- Conditional three years performance and value at risk by Rating.

Performance (t+36)	Stars rating (t)				
	1-star	2 stars	3-stars	4-stars	5-stars
Sharpe_	0.699	0.721	0.748	0.790	0.832
Return_	15.254	15.252	15.296	15.837	16.403
Alpha_	-4.649	-2.519	-2.327	-1.173	-0.660
VaR99_	7.116	6.691	6.278	6.626	6.211
Performance (t+36)	**Analyst rating (t)**				
	Negative	Neutral	Bronze	Silver	Gold
Sharpe_	-	0.823	0.781	0.740	0.787
Return_	-	16.801	15.730	15.005	15.794
Alpha_	-	-1.687	-1.954	-1.684	-0.825
VaR99_	-	7.170	6.369	6.080	6.385

Note: Sharpe is the Sharpe Ratio calculated in a three-annual basis, Alpha is the beta-adjusted return over a three-year period, Return is the net 3 years return. VaR99 is the value at risk at a confidence level of 99%.

4.2. Models

In the section, we analyse if the analyst rating can help to identify products that will outperform their peers in the next period after the initial rating. At the same time, we also include separately stars ratings to compare both alternatives, taking into account that investors take their decisions based on backward looking ratings. Finally, we use both ratings to evaluate if taking decisions combining good stars and analyst mutual funds can help in the selection of outperformers.

Firstly, we estimate some models to evaluate out of sample performance based exclusively on the analyst rating and their main pillars. If analyst rating is a forward-looking measure that reflects the expectations of analyst about future performance in the long run, we expect that higher ratings will obtain higher future performance. We calculate out of sample risk adjusted returns for 12 and 36 months after the initiation rating and then we regress the different metrics using Gold, Silver and Bronze indicators variables. For 12 months models, we estimate a panel data regression model (random effects) and for 36 months we use OLS regression because we only have one period of three years after the initial grade. The methodology based on panel

data can control for individual effects with advantages like the reduction of collinearity and efficiency, among others (Baltagi, 2010). The following equations are estimated:

$$Y_{i,t+k} = \alpha_i + \beta_1 Gold_{it} + \beta_2 Silver_{it} + \beta_3 Bronze_{it} + \sum_i \text{Month}_t + \sum_j Category_j + \varepsilon_{it}$$

$$Y_{i,t+k} = \alpha_i + \beta_1 Gold_{it} + \beta_2 Silver_{it} + \beta_3 Bronze_{it} + \beta_4 Netexp_{it} + \beta_5 logSize_{it} + \beta_6 ManagerTen_{it} + \sum_i \text{Month}_t + \sum_j Category_j + \varepsilon_{it}$$

Where $Y_{i,t+k}$ is the performance obtained by the fund i for 12 of 36 months after the initial rating. As performance measures we use Alpha, Sharpe and Total Return. Gold, Silver and Bronze are indicator variables considered as recommended funds, that take the value of 1 when the mutual fund is rated as one of these categories and 0, otherwise. Category are dummy variables for the nine categories considered in the study. Finally, α_i and β_1,β_2,β_3, β_4, β_5, and β_6 are parameters of the regression and ε_i the term error.

We estimate the models first controlling by month and categories, and then including some additional controls like expenses, size and age. Following Chen and Huang (2011) and Armstrong *et al.* (2016), we also include as control variables the manager tenure (ManagerTen) of the fund, the costs measured by the net expenses ratio (Netexp) and the size of the mutual fund (LogSize).

4.3. Results for Analyst ratings (forward looking)

Table 17 shows the results of regressing the risk adjusted return with the analyst indicators variables after twelve months since the rating is available, through a robust random effects panel data model. As in previous works, the models were initially estimated without control variables to evaluate the effect of selecting funds based exclusively on the rating analyst. Subsequently, other explanatory variables such as costs, size and management experience have been included. In all cases, category and time control variables have been included. As you can see, the results depend on the metric used, but in general, only funds classified as gold show better 12-month performance than the "Not recommended" funds in terms of Sharpe's ratio. Over 36 months the results are more disappointing and the only significant sign is negative. Therefore, our analysis reveals the inability of analyst ratings to identify funds that outperform their peers except in the case of gold funds, where the results show a better performance than those classified as not recommended. The differences with previous studies may be due to the fact of considering in our study a different sample focused exclusively on US, the panel data methodology and the different period considered.

Table 17- Analyst ratings and out-of-sample performance after 12 and 36 months

Variable	Alpha-12	Sharpe-12	AlphaC-12	SharpeC-12	Alpha-36	Sharpe-36	AlphaC-36	SharpeC-36
Gold	0.7204	0.2023***	0.7288	0.3572***	0.311	0.0368	0.024	-0.012
Silver	-0.1788	0.0219	-0.2108	0.1581	-0.3122	-0.0945*	-0.1202	-0.0711
Bronze	-0.1094	-0.0334	0.4072	0.1216	-0.1893	-0.047	0.1439	-0.0055
Netexp_	-	-	-0.7281	-0.0824	-	-	-0.8025	-0.1673***
logSize	-	-	0.2006	-0.0123	-	-	0.4997***	0.0520***
ManagerTen_	-	-	-0.0118***	-0.0039***	-	-	-0.0095**	-0.0009**
USlargeblend	-1.8096***	0.2660***	-2.7008***	0.2364***	-2.2694***	0.1259*	-2.8618***	0.0785
USlargegrowth	-2.6385***	0.1472**	-2.0243***	0.2421***	-0.9392	0.2351***	-1.3944**	0.1900***
USlargevalue	-1.2823***	0.2819***	-1.5891***	0.1575	-1.2142**	0.1176*	-2.0041***	0.037
USmidcap	-2.1359***	0.1582**	-1.9125***	0.2656***	-2.0007***	0.1896**	-2.3070***	0.1563**
_cons	0.6169	0.9160***	-1.7682	1.6383***	0.8999	1.1514***	-8.1729**	0.2718
N	790	852	529	562	172	182	161	170
r2	0.0003	0.007	0.0756	0.4327	0.1203	0.1195	0.2466	0.2983

This table reports the coefficients for Panel Data models for the Alpha and Sharpe performance measures after 12 and 36 months. Gold, Silver and Bronze are dummies to control the medal of a fund obtained from Morningstar Analyst Rating. ManagerTen, the manager tenure, Netexp, the net expense ratio and LogSize, the logarithm of the assets of the fund, are control variables. Finally, USlargeblend, USlargegrowth, USlargevalue and USmidcap are dummies to control for categories. N is the number of observations, r2 is a measure of the goodness of fit. *Significant at 10%; ** significant at 5% and * significant at 1%.**

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

4.4. Results for some Pillars included in the Analyst ratings (forward looking)

In this section, we have included the pillars in which the analyst ratings are broken down to analyse whether the future performance is related to any specific dimension. Specifically, we consider only three pillars, excluding Price, because the expense ratio and performance are already included as a control variable, since the quantitative ratings that contain such information are subsequently analysed. The obtained results show, in general, non-significant coefficients, with both negative and positive sign. The fact that there are very few mutual funds with a negative pillar leads to a comparison between fundamentally positive and neutral pillars. The high subjectivity that can lead to the evaluation process and the difficulty to establish a threshold between a positive and a neutral evaluation can explain the results obtained.

Table 18- Pillars and out-of-sample performance after 12 and 36 months

Variable	Alpha 12	Sharpe 12	Alpha C-12	Sharpe C-12	Alpha 36	Sharpe 36	Alpha C-36	Sharpe C-36
Parentpos	0.5001	0.0688*	-0.0211	-0.0089	0.21	0.0294	-0.1735	-0.0307
Peoplepos	-0.543	-0.0881	-0.1532	-0.0304	0.1531	0.0431	-0.0274	0.007
Processpos	-0.284	-0.0275	0.3184	0.0293	-0.4298	-0.1147*	0.4842	0.0356
USlargeblend	-2.1777***	0.2538***	-2.8730***	0.1849***	-2.5693***	0.0906	-3.2014***	0.0239
USlargegrowth	-1.8044***	0.2255***	-1.9416***	0.2335***	-1.1939	0.1937**	-1.4232**	0.1829**
USlargevalue	-1.1168*	0.2754***	-1.7632***	0.1844**	-1.2380*	0.1343	-2.2574***	0.0195
USmidcap	-1.5612**	0.2857***	-1.6983**	0.2659***	-1.7527**	0.1782*	-1.9586**	0.1650*
Netexp_	-	-	-0.9361*	-0.1591***	-	-	-0.8113	-0.1491**
logSize	-	-	0.4038**	0.0470**	-	-	0.6220***	0.0736***
ManagerTen_	-	-	-0.0088**	-0.0009**	-	-	-0.0095*	-0.0010*
_cons	3.3915***	0.3681***	0	0	1.1183	1.1884***	-11.0519**	-0.2456
N	527	565	478	493	130	130	123	123
r2	0.1586	0.8135	0.1647	0.8243	0.1219	0.0974	0.2539	0.3118

This table reports the coefficients for Panel Data models for the Alpha and Sharpe performance measures after 12 and 36 months. Parentpos, Peoplepos, Processpos are dummies which take the value 1 in case the Pillars Parent, People and Process of the Morningstar Analyst Rating is positive, 0 otherwise. ManagerTen, the manager tenure, Netexp, the net expense ratio and LogSize, the logarithm of the assets of the fund, are control variables. Finally, USlargeblend, USlargegrowth, USlargevalue and USmidcap are dummies to control for categories. N is the number of observations, r2 is a measure of the goodness of fit of the model. *Significant at 10%; ** significant at 5% and * significant at 1%.**

4.5. Results for stars ratings (backward looking)

The models estimated for the stars rating display different results than the analyst, showing that the ratings of 4 outperform those considered as not recommended according to this criterion in terms of alpha and Sharpe but only in the short term (12 months) and without control variables. Given that four one-year periods are used in the panel regression, the results indicate that investing for the term of one year starting at any one of them, yields better results for investors with four stars funds. However, this superiority is not maintained in the long term,

suggesting, as for analyst, the need to monitor the portfolio on a yearly basis and to verify that the funds are preferably rated 4 stars.

Table 19- Stars ratings and out-of-sample performance after 12 and 36 months

Variable	Alpha 12	Sharpe 12	Alpha C-12	Sharpe C-12	Alpha 36	Sharpe 36	Alpha C-36	Sharpe C-36
stars5	0.1855	0.0404	-0.3248	0.034	-0.2197	0.0227	0.1997	0.0427
stars4	0.9753**	0.0975*	0.0327	0.0544	-0.2543	-0.0016	-0.2489	-0.0073
stars3	0.1446	0.0221	-0.701	0.0166	0.2108	0.0125	0.7859	0.0703
Netexp_	-	-	-0.6308	-0.1121**	-	-	-0.4743	-0.1342***
LogSize_	-	-	0.2432	0.0342*	-	-	0.5132***	0.0500***
ManagerTen_	-	-	-0.0076**	-0.0009**	-	-	-0.0085**	-0.0008**
USlargeblend	-1.6681***	0.2541***	-2.8737***	0.1457**	-2.3813***	0.1524***	-3.3087***	0.0652
USlargegrowth	-2.3913***	0.1332***	-1.9423***	0.1703***	-1.4986***	0.2284***	-2.1084***	0.1770***
USlargevalue	-1.1255**	0.2728***	-1.5445***	0.1567**	-1.5083***	0.1261**	-2.3011***	0.0473
USmidcap	-1.9233***	0.1812***	-1.9599***	0.2064***	-2.1067***	0.2036***	-2.5243***	0.1639**
_cons	-1.1438**	0.3607***	-3.2438	0	1.2343**	1.1152***	-8.5187**	0.2432
N	873	930	587	618	218	229	201	211
r2	0.1334	0.8217	0.1747	0.841	0.1238	0.0964	0.2734	0.273

This table reports the coefficients for Panel Data models for the Alpha and Sharpe performance measures after 12 and 36 months. Stars5, stars4, stars3 are dummies which reflect the number of stars of Morningstar Star Rating. ManagerTen, the manager tenure, Netexp, the net expense ratio and LogSize, the logarithm of the assets of the fund, are control variables. Finally, USlargeblend, USlargegrowth, USlargevalue and USmidcap are dummies to control for categories. N is the number of observations, r2 is a measure of the goodness of fit of the model. *Significant at 10%; ** significant at 5% and * significant at 1%.**

4.6. Results for best stars and analyst ratings (combining forward and backward looking)

In this section, we want to check if the combination of both ratings can help in the process of identifying outperformers. Thus, we created different indicator variables that result from the combination of the best analyst (Gold to Bronze) and stars ratings (4 or 5 stars) and considering the rest as "Not recommended". The results of the Table 20 show that within 12 months only the gold funds of four or five stars outperform the funds with the worst ratings. When we take the 36-month term, the differences are positive in terms of Sharpe ratio for the categories bronze4or5, gold3stars and bronze3stars. The rest of the signs are generally positive, except in the case of silver3stars, but not significant. The results of the joint analysis of the funds show that when the ratings are used in isolation, only the four- or five-star funds and gold funds outperform the poorer quantitative rating in the short term. When both criteria are combined, again, the funds that result from combining gold and four or five stars perform better for the 12-month term, but not the 36-year term. This means that investors who base their decisions on both criteria must monitor portfolios annually and check that they continue to maintain both ratings. On the other hand, the combination of the two ratings does have medium-term differentiation results, with a higher performance in terms of Sharpe's ratio for bronze (3, 4 or 5 stars) or three-star gold funds, presenting only a negative sign the three-star silver backgrounds.

Table 20- Stars and analyst rating combination and out-of-sample performance after 12 and 36 months.

Variable	Alpha 12	Sharpe 12	Alpha C-12	Sharpe C-12	Alpha 36	Sharpe 36	Alpha C-36	Sharpe C-36
gold4or5	0.9131*	0.0970*	0.1859	0.0182	-0.0055	0.1124	-0.7422	-0.0224
silver4or5	0.288	0.0325	-0.6254	-0.0328	-0.4702	0.081	-0.711	0.0289
bronze4or5	0.0651	0.0369	0.0327	0.0469	-0.0156	0.1152**	0.0466	0.0598
gold3stars	0.3567	0.1063	-0.5773	-0.0386	1.2412*	0.3196**	1.0858	0.2475*
silver3stars	-0.445	-0.0272	-1.0164	-0.02	-0.4032	-0.1295	-0.5996	-0.1546*
bronze3stars	-0.1211	-0.0208	-0.799	-0.0341	1.1077	0.1938**	0.9349	0.1988**
USlargeblend	-1.6695***	0.2637***	-2.9036***	0.1426**	-2.9453***	0.5724***	-3.5135***	0.4660***
USlargegrowth	-2.3087***	0.1507***	-2.0591***	0.1648**	-2.0367***	0.6547***	-2.4280***	0.5690***
USlargevalue	-1.0088**	0.2855***	-1.5626***	0.1584**	-1.9786***	0.5309***	-2.3603***	0.4557***
USmidcap	-1.8791***	0.2019***	-2.0078***	0.2109***	-2.2934***	0.6345***	-2.5141***	0.5770***
Netexp_	-	-	-0.6264	-0.1216**	-	-	0.0439	-0.0633
logSize	-	-	0.2281	0.0350*	-	-	0.4279***	0.0520**
ManagerTen_	-	-	-0.0070**	-0.0008**	-	-	-0.0062*	-0.0002
_cons	-0.8995*	0.3755***	0	0	1.5764***	0.6344***	-6.8484**	-0.3307
N	874	942	588	627	331	484	305	406
r2	0.1254	0.8213	0.1715	0.8425	0.1979	0.3566	0.2975	0.3842

This table reports the coefficients for Panel Data models for the Alpha and Sharpe performance measures after 12 and 36 months. Gold4or5 is a dummy variable with the value of 1 if the fund is rated Gold with 4 or 5 stars, silver4or5 is a dummy variable with the value of 1 if the fund is rated Silver with 4 or 5 stars, bronze4or5 is a dummy variable with the value of 1 if the fund is rated Bronze with 4 or 5 stars, gold3stars is a dummy variable with the value of 1 if the fund is rated Gold with 3 stars and silver3stars is a dummy variable with the value of 1 if the fund is rated Silver with 3 stars. ManagerTen, the manager tenure, Netexp, the net expense ratio and LogSize, the logarithm of the assets of the fund, are control variables. Finally, USlargeblend, USlargegrowth, USlargevalue and USmidcap are dummies to control for categories. N is the number of observations, r2 is a measure of the goodness of fit of the model. *Significant at 10%; ** significant at 5% and * significant at 1%.**

5. Robustness

We conducted some additional robustness tests to check the consistency of our results and to provide other complementary analysis. In addition to panel data we also use the quantile regression to extend the regression model to conditional quantiles of the different performance metrics because it is more appropriate for a heterogeneous mutual fund universe where strategies and objectives can vary (Chen and Huang ;2011). This model let us capture information about the coefficients at different quantiles of the dependent variable given the set of endogenous variables (star rating). In addition, the conditional quantile regression developed by Koenker and Bassett (1978) deals well with skewed distributions of fund performance. In particular, we adopt the bootstrapping method proposed by Efron (1979) and implemented in the software Stata 12.

Given Y_i as the different performance metrics used in this paper, and X_i as a vector of exogenous variables representing the rating of the fund, the quantile model can be written as:

$$y_i = X_i' \beta_\phi + u_{\phi i}$$

Assuming that:

$$Quant_{\phi}(y_i|X_i) = X_i'\beta_{\phi}$$

$$Quant_{\phi}(u_{\phi i}|X_i)=0$$

As can be seen in Table 21, the results of the quantile regression show that in general the signs are not significant for most of the ratings, with gold being the only ones that outperform those not recommended in quartile 25 and 75. When included the management costs, the gold ratings that are significant, cease to be, and it is therefore reasonable to think that both the size of the funds and the costs can explain the differences in performance more than analyst ratings.

Table 21- Quantile regression (Out of sample performance after 12 months).

Variable	Alpha-12	Sharpe-12	AlphaC-12	SharpeC-12
q25				
Gold	0.7348	0.1542**	-0.7879	0.0494
Silver	-1.3122**	-0.0175	-1.4756**	-0.0429
Bronze	-0.3791	0.0083	-0.0466	0.0386
Netexp_	-	-	-2.6220***	-0.1997***
logSize	-	-	0.2811*	0.0165
ManagerTen	-	-	-0.0056	-0.0008
cons	-3.9877***	0.0974	-4.4653	1.2791**
q50				
Gold	0.1641	0.0577	-0.4749	0.0068
Silver	-0.3934	0.0066	0.066	0.0096
Bronze	-0.1826	-0.0113	0.3432	0.0671
Netexp_	-	-	-0.4713	-0.1107
logSize	-	-	0.4343***	0.0399
ManagerTen	-	-	-0.0104**	-0.001
_cons	-0.3282	0.4845***	-7.6554**	1.2672*
q75				
Gold	1.1570**	0.0809	0.8035	-0.0494
Silver	1.1729**	0.0853	1.0237	0.0723
Bronze	0.4998	0.0378	0.6718	0.0846*
Netexp_	-	-	-0.2892	-0.1315
logSize	-	-	0.1479	0.0543**
ManagerTen	-	-	-0.0025	-0.0014***
_cons	2.4698***	0.6826***	-0.6839	1.5432**
N	790	852	529	562

This table reports the coefficients for Quantile regression. Alpha12 is the beta-adjusted return over a one-year period; Sharpe12 is the yearly risk-adjusted return. Gold, Silver and Bronze are dummies to control the medal of a fund obtained from Morningstar Analyst Rating. ManagerTen, the manager tenure, Netexp, the net expense ratio and LogSize, the logarithm of the assets of the fund, are control variables. N is the number of observations. *Significant at 10%; ** significant at 5% and * significant at 1%.**

One reason that might explain why analyst ratings were not very significant in previous analysis may be because "Morningstar analyst rating is a qualitative, forward-looking measure that reflects the analyst's expectation of the future performance relative to its peers over a

business cycle" and thus the analysis is only coherent when we compare peers, or the same categories. Thus, in this section we try to make the analysis only for large-scale categories because they have enough number of mutual funds. As can be seen in Table 22, the analysis by categories shows that gold funds outperform neutral or negative funds in the 12-month term and partially in the 36-month term in the case of the large-blend category. The same happens in the Growth category, where gold funds return again outperform to those not recommended in most of the metrics, terms and including control variables. This capacity is not observed in the case of Value, where most of the signs are positive but not significant. Therefore, the gold ratings allow identify funds that exceed their peers, mainly for a term of one year but not in all categories. Comparing the results with the main model we observe that in general for some categories we obtain that gold ratings are more significant but in general, very few differences has been obtained.

Table 22- Analysis for large-scale categories.

Category Blend								
Variable	**Alpha-12**	**Sharpe-12**	**AlphaC-12**	**SharpeC-12**	**Alpha-36**	**Sharpe-36**	**AlphaC-36**	**SharpeC-36**
Gold	1.1925*	0.2853**	2.1843**	0.5380***	0.7194	-0.0231	1.6771**	0.0472
Silver	-0.4759	0.0583	-0.6327	0.0966	-0.0169	-0.1272	0.0149	-0.1131
Bronze	-0.3932	0.0063	-0.0636	0.078	-0.1105	-0.0896	0.8181	-0.0028
Netexp_	-	-	-0.4452	0.0028	-	-	-0.9610*	-0.1046*
logSize	-	-	-0.3183	-0.0476	-	-	-0.0095	0.0165
ManagerTen	-	-	-0.0218***	-0.0055***	-	-	-0.0240***	-0.0025***
_cons	-1.1341*	1.1393***	7.6801*	2.7090***	-1.7539***	1.1191***	1.2602	1.2444***
N	191	215	135	146	90	106	39	46
r2	0.0187	0.0291	0.1054	0.3565	0.0277	0.0283	0.4702	0.4784
Category Growth								
Variable	**Alpha-12**	**Sharpe-12**	**AlphaC-12**	**SharpeC-12**	**Alpha-36**	**Sharpe-36**	**AlphaC-36**	**SharpeC-36**
Gold	2.5081***	0.2483	2.9677**	0.4703**	2.5668**	0.1345	3.0034**	0.2017**
Silver	0.1032	-0.2511*	-0.9034	-0.1225	-0.3255	-0.1021	-0.274	-0.0001
Bronze	0.7343	0.0338	0.2825	0.1469	0.3727	0.0085	0.1689	0.0036
Netexp_	-	-	0.0522	0.1522	-	-	3.7374	0.4183**
logSize	-	-	0.3867	-0.0268	-	-	0.8891**	0.0960***
ManagerTen	-	-	-0.0053	-0.0026	-	-	-0.0071	-0.0010*
_cons	-2.5865***	1.0803***	-9.2674	1.8833*	-1.9022***	1.0665***	-23.0246**	-1.0813
N	166	171	101	103	77	77	29	29
r2	0.0034	0.0521	0.0138	0.2426	0.1324	0.0367	0.4412	0.5076
Category Value								
Variable	**Alpha-12**	**Sharpe-12**	**AlphaC-12**	**SharpeC-12**	**Alpha-36**	**Sharpe-36**	**AlphaC-36**	**SharpeC-36**
Gold	0.9163	0.1868	0.818	0.1301	0.1344	-0.0405	0.4533	-0.0762
Silver	-0.2367	0.1761	0.0135	0.4033	-0.1647	-0.0641	0.22	-0.0588
Bronze	-0.5388	-0.1056	0.9535	0.5165	-0.4786	-0.1494	0.8571	-0.0562
Netexp_	-	-	-3.1108***	-0.2363	-	-	-2.4342***	-0.1074
logSize	-	-	0.2379	0.0821	-	-	0.1321	0.036
ManagerTen	-	-	-0.0032	-0.0098**	-	-	-0.0156	-0.0007
_cons	-0.4915	1.1819***	-3.2229	0.0024	-0.694	1.1116***	-0.6743	0.6336
N	104	114	71	77	50	55	23	26
r2	0.0043	0.0456	0.0137	0.371	0.0132	0.0386	0.3171	0.1697

This table reports the coefficients for Panel Data models for the Alpha and Sharpe performance measures after 12 and 36 months large-blend, large growth and large value categories. ManagerTen, the manager tenure, Netexp, the net expense ratio and LogSize, the logarithm of the assets of the fund, are control variables. N is the number of observations, r2 is a measure of the goodness of fit of the model. *Significant at 10%; ** significant at 5% and * significant at 1%.**

6. Conclusions

Many investors select their investments in mutual funds based on quantitative rating. However, the selection of funds based exclusively on quantitative ratings, excludes a set of qualitative factors that can explain future performance. Morningstar has two systems to classify mutual funds based on qualitative aspects: Morningstar Analyst Rating and Morningstar Stewardship Grade. Morningstar Analyst Ratings are forward-looking qualitative and quantitative analyses of mutual fund about five pillars: Process, Performance, People, Parent and Price, that includes factor like the cost, past performance, quality of management, interest alignment, etc. Morningstar Analyst Ratings are based on the convictions that funds will outperform their benchmarks over the long term. Despite there exists several studies about quantitative ratings, Analyst Ratings is not as popular and very few research has been done about it. Kamal (2013) and Armstrong, *et al.* (2016) are the unique authors that focus their research on Analyst ratings.

In this paper, we assess to what extent selecting mutual funds based on Morningstar Analyst and Star ratings criteria has an impact on the performance of investors. In particular, we endeavour to answer if good analyst ratings outperform non-recommended ones in the short (12 month) and long term (36 month) and if it is useful to combine both ratings in the screening process to identify good future performers. The data has been collected from Morningstar Direct database covering the period August 2012 to August 2016. We selected US equity funds with the previously mentioned ratings. Our conclusions support the ability of Gold ratings to select funds that will behave better in terms of future performance. Our results are in line with previous empirical evidence found in Kamal (2013) and Armstrong *et al.* (2016) but we have found little evidence that, on average, funds with a better Analyst Rating (Gold) have a better performance in terms of risk adjusted measures (alpha and Sharpe). The predictability is observed in several analyses done in one year ahead but not for three-years. This evidence is more relevant in the case of the analysis made by investment style´s category. In the analysis of if the pillars in which the analyst ratings are broken down we do not find evidence that future performance is related to any of these specific dimensions. The combination of Stars and Analyst ratings does have medium-term differentiation results, with a higher performance in terms of Sharpe's ratio for bronze (3, 4 or 5 stars) or three-star gold funds.

The inclusion of other variables such as costs, size and manager tenure reflects the importance of considering other variables for fund´s selection. Nevertheless, in several estimations Gold ratings are still significant in explaining performance, indicating that costs are not the only factor that determines the predictive power of qualitative ratings. Our results support

the use of qualitative ratings in the investment funds selection process, accompanied by other variables.

Chapter IV: Does Sustainability Score Impact Mutual Fund Performance?

1. Introduction

Socially Responsible Investment (SRI) also known as sustainable, responsible and impact investing is "an investment discipline that considers environmental, social and corporate governance (ESG) criteria to generate long-term competitive financial returns and positive societal impact" (US SIF, n.d.[3]). Other names for SRI are social investment, sustainable, socially conscious, green, responsible or ethical investing. These terms summarize any investment strategy which searches for a financial return and encourages corporate practices that promote environmental care, consumer protection and human rights. So, SRI is a "generic term covering sustainable, responsible, socially responsible, ethical, environmental, social investments and any other investment process that incorporates environmental, social and governance issues" (GSIA, 2017). According to 2016 Global Sustainable Investment Review[4] (GSIA, 2017) in 2016 there are $22.89 trillion assets being professionally managed under SRI strategies in the world, an increase of 25% since 2014, therefore sustainable investing constitutes a major force across global financial markets. This report highlights the fact that Europe accounts for over half the global SRI professionally managed assets (52.6%), and the United States more than a third (38.1%). In Europe, total professionally managed assets committed to SRI strategies grew by 11.7% from 2014 to 2016 to reach $12.04 trillion, but this growth is much lower than what is happening in other areas, in particular Japan.

The growing interest in SRI in recent years has led to several organizations assessing mutual funds on how well the underlying companies perform on environmental, social, and governance (ESG) issues. In 2016, two initiatives were launched to provide investors with independent assessments on ESG issues. Morningstar launched a Morningstar Sustainability Rating to assess about 20,000 mutual funds and exchange-traded funds (ETFs) around the

[3] See http://www.ussif.org. US SFI is a Forum for Sustainable and Responsible Investment. Legally is a United States-based membership association. Members of the association are professional investors such investment management and advisory firms, mutual fund companies, research firms, financial planners and advisors, pension funds, etc.

[4] This is the third edition of the biennial *Global Sustainable Investment Review* that presents results from Europe, United States, Canada, Asia, Japan, and Australia and New Zealand. This Review is publishes by Global Sustainable Investment Alliance (GISA), which members are US SFI in the United States, Eurosif in Europe, etc.

world. The idea of the Morningstar Sustainability Rating is classifying mutual funds about ESG factors relative to their Morningstar category peers. The other initiative was that MSCI launched MSCI ESG Fund Metrics to measure over 100 scores and metrics of ESG characteristics of the portfolio holdings of approximately 21,000 mutual funds and ETFs[5]. The advantage of these products is that it makes it possible to find sustainable funds even if they aren't labelling themselves specifically as funds that support a SRI approach. The use of these scores shows an important difference with previous studies, which compare SRI funds with an index or the most advanced studies apply, a so called matching approach, i.e. they compare the performance of SRI and non-SRI investment funds with similar characteristics (fund size, fund age, etc.) to properly considered management and transaction costs for both SRI funds and the conventional funds (see Mallin, Saadouni and Briston, 1995; Gregory, Matatko and Luther, 1997; Statman, 2000; Kreander, Gray, Power and Sinclair, 2002, 2005; among others).

One important research question in the mutual fund industry about SRI investing is to know how do SRI mutual funds perform? There are several studies that have demonstrated that companies with social responsibility policies and practices are good investments. For example, a recent paper of Friede, Busch and Bassen (2015) conducted a meta-analysis of about 2,200 empirical studies, making it the most comprehensive review of academic research on this topic. They found that the majority of studies show a positive correlation between ESG factors and financial performance. But even despite the investigations carried out to date there is still a debate about whether these types of investments can create value for investors or not and why they put their money here. Although according to Lewis and Mackenzie (2000) and Webley, Lewis and Mackenzie (2001) some investors in SRI funds are willing to accept lower returns for their moral stance, the performance of SRI funds and conventional funds is still an open question. As Junkus and Berry (2015) sustain, after a review of the most recent work in major finance journals on socially responsible investment (SRI), "the performance of SR mutual funds and indexes are not generally significantly different to conventional funds or indexes, but again these results are also highly dependent on model specification, time period, benchmark, and other characteristics of the study".

Authors such as Luther, Matatko and Corner (1992) and Mallin, Saadouni and Briston (1995) support the idea that SRI funds outperform market indexes. But the more

[5] See https://www.msci.com/esg-fund-metrics.

conventional theory is that SRI mutual funds have the same return like any other funds, and authors as Hamilton, Jo and Statman (1993), Kreander, Gray, Power and Sinclair (2002, 2005), Gregory and Whitakker (2007) and Bauer, Derwall and Otten (2007) or Humphrey, Warren and Boon (2016) are in line with this theory. Another theory defends that choosing SRI funds are basically a "trade off" between investing in SRI and returns, so SRI investments underperform the benchmark (e.g. White, 1995). One important recent paper is Nofsinger and Varma (2014), that provides a new perspective on the eternal question of whether ESG criteria are financially material or not. The authors find that risk adjusted performance is different between socially responsible (SR) and conventional mutual funds, and this difference depends on the state of the market. SR mutual funds outperform conventional mutual funds during periods of market crisis, but in non□crisis SR funds underperform conventional funds.

Previous research has studied the effect of sustainability on performance using exclusively a dichotomous variable to differentiate between socially responsible funds and conventional funds. However, the results could be biased because under socially responsible they could have funds with very different levels of sustainability and, at the same time, it could also appear that the differences with "non-responsible" are also not very relevant. Statman and Glushkov (2016) conclude that there is a lack of clearly defined criteria to distinguish mutual funds as socially responsible results in inconsistently applied classifications that makes it difficult to measure the performance of socially responsible investments. Traditional methodology in the empirical research is benchmarking with indices or most recently matched pair analysis, which was initially applied by Maillin, Saadouni and Briston (1995) and is based on comparing returns of SRI funds and conventional funds with similar characteristics in terms of volume of assets and interception dates. For this reason, the inclusion of sustainability scores in our work can allow us to evaluate whether the degree of sustainability of the portfolio in which the funds are invested has a positive effect on performance. As far as we know, only El Ghoul and Karoui (2017) use a CSR scores to study the effect on fund performance and flows, concluding that higher values display poorer performance and weaker performance-flow relation. From an investor point of view, the advantage of using sustainability scores is that they can select their SRI taking into consideration the funds with better scores, whether or not they are declared as an SRI fund[6].

[6] SRI funds and conventional funds have a different industry asset composition, so assets of SRI funds are highly concentrated within certain industries (renewable industries, etc.) while industries such as mining, oil

This paper adds to the growing literature on SRI by specifically examining the effect of the degree of sustainability, measured though Morningstar Sustainability scores included in Morningstar Direct in 2016. In particular, we assess the effect of sustainability scores and the different dimensions in which the score is subdivided (Environmental, Social and Governance) in the performance, in addition to downside risk and the flow of funds. On the other hand, the conventional dichotomous variable has been added to the models to evaluate to what extent the results may differ. Our empirical evidence also contributes to the literature on mutual funds that discusses whether applying a particular investment screening in portfolio selection affects the mutual fund performance (see, e.g., Bauer, Derwall and Otten, 2007 or Muñoz, Vargas and Marco, 2014). The Portfolio theory argues that narrowing the universe of assets restricts diversification opportunities and thus the risk-adjusted performance (Rudd, 1981); whereas an opposing view sustains (Hill, Ainscough, Shank and Manullang, 2007 or Chegut, Schenk, and Scholtens, 2011) that restricting investment screening allows the identification of companies with higher growth potential and better management, therefore leading to a better financial performance and risk profile. Sustainable mutual funds apply a specific portfolio screening by concentrating investments in socially conscious businesses. Although there is profuse empirical literature on the impact of social responsibility of the performance, little is known about the screening based regarding the level of sustainability.

Our empirical results show that a large number of funds are not declared sustainable but their portfolio is comparable to sustainable mutual funds. Furthermore, Sustainability score is significant in explaining the level of performance, downside risk and flows. We also achieved equivalent results for the three dimensions of sustainability (environmental, social and corporate). The signs are different on performance and downside risk when the conventional dummy to declare social mutual funds is used.

The remainder of this paper is laid out as follows. In Section 2, we review the related literature on SRI performance; in Section 3 we describe our data and the performance evaluation metrics; in Section 4 we describe our empirical methods and results and; finally, in Section 5 we draw conclusions from our research.

and gas are underweighted. Benson, Brailsford and Humphrey (2006) concluded that SRI funds have a different industry composition and that this affected performance.

2. Literature Review

Over the last few years, SRI investment research has been growing. CFA Institute, which is a global association for investment professionals, states that "a key idea in the discussion of ESG issues is that systematically considering ESG issues will likely lead to more complete analyses and better-informed investment decisions" and "... that every investment analyst should be able to identify and properly evaluate investment risks, and ESG issues are a part of this evaluation" (CFA Institute, 2015). For this association, there are basically two investors interested in considering ESG issues: *value-motivated* and *values-motivated investors*. We focus on the first kind of investors who are concerned about the financial performance of their SRI funds.

Hamilton, Jo and Statman (1993) developed three hypotheses regarding the performance of SRI mutual funds. The first hypothesis is that SRI fund performance equals that of conventional funds, which is consistent with a market that does not regard the social responsibility feature. Second hypothesis, SRI fund performance is lower than that of conventional funds, which is consistent with a market that values the social responsibility feature. Finally, third hypothesis is that SRI fund performance is higher than that of conventional funds. There are several arguments which could explain why SRI mutual funds can outperform in financial terms the conventional funds (which do not consider ESG factors). First, SRI mutual funds have a higher proportion of their portfolio in the segment of small companies; these companies are better adapted to market changes (Luther, Matatko and Corner, 1992; Gregory, Matatko and Luther, 1997) and also may be more profitable in the long run. Second, social companies are more efficient, better managed and develop better in the market (Hamilton, Jo and Statman, 1993). From a theoretical point of view (see a discussion in e.g. Margolis, Elfenbein and Walsh, 2009 or Flammer, 2015). social companies can reduce costs (penalties, etc.) or increase revenues (innovative products, greater employee effort, better public perception, increasing the likelihood that consumers will purchase the company's products or its share price, attract socially conscious customers, etc.). In contrast, one important argument of the detractors of SRI funds is that the universe of possible investments of these funds (individual companies) is small, so they assume a higher investment risk because the lack of diversity (Chegut, Schenk and Scholtens, 2011). But authors such as Kurtz (1997) or Goldreyer and Diltz (1999) argue that SRI mutual funds managers, need more information than conventional funds about the companies in which they invest, they base their decisions on a deeper, more complete and higher quality

information, resulting in a significant reduction in the risk of their investment decisions. Empirical evidence of some authors, such as Luther, Matatko and Corner (1992) and Maillin, Saadouni and Briston (1995) support the idea that SRI funds outperform conventional investment. But, there is also evidence to support that social responsibility investments are neutral to financial performance (Hamilton, Jo and Statman, 1993; Kreander, Gray, Power and Sinclair, 2005; Gregory and Whittaker, 2007; Bauer, Derwall and Otten, 2007; among others), or that SRI funds underperform conventional investment (e.g. White, 1995).

The first study about SRI investment was done for Luther, Matatko and Corner (1992) where these authors found that SRI investment funds did not under or out-perform the index benchmark. They used 15 British Ethical funds finding weak evidence that 15 UK SRI fund outperformed two stock market indices. Hamilton, Jo and Statman (1993) conducted a similar study where the difference of means of excess returns was not significant and only one of 17 mutual funds had a positive Jensen`s Alpha. Luther and Matatko (1994) improved their prior work by including a small market index and they concluded that the excess returns of SRI funds are strongly influenced by the low capitalization of the small cap stocks. The study also shows that SRI funds have a neutral effect on performance. White (1995) researches US and German mutual funds using a simple regression against an environmental market index, showing that the SRI investments underperform the benchmark in terms of three performance measures (Jensen`s Alpha, Treynor Ratio and Sharpe). In this research, the author used a sample of 6 US funds and 5 German SRI Investment funds.

All previous studies used benchmark as an index, so they have the problem of what is the appropriate index. Mallin, Saadouni and Briston (1995) avoided this problem by using a matched pair analysis to compare SRI mutual funds and conventional funds in UK. The authors matched 29 SRI mutual funds to conventional ones using as criteria, the size and the age. Their results showed no differences in the performance of both samples using the performance measures Sharpe and Treynor, but they found that ethical funds did better than the non-ethical funds when the Jensen performance measure was used. Gregory, Matatko and Luther (1997) studied 18 SRI funds where the investment area and the fund type were considered. They did not find differences in performance against conventional funds. Statman (2000) studied the performance of 31 US SRI mutual funds and the Domini 400 Social-Index (DSI) from 1990 to 1998. The results show that only some SRI funds could underperform the benchmark (S&P 500 or DSI). But in general, SRI funds obtained a

similar performance to S&P 500, DSI and conventional funds. Kreander, Gray, Power and Sinclair (2002) using a matching procedure and the age, size, country and investment universe of the fund as variables. The study included mutual funds from Sweden, Netherlands, Norway, Germany, UK and Switzerland, and as performance metrics the Sharpe, Jensen`s Alpha and Treynor Ratio. Their results showed that SRI funds' performance was very similar to those of conventional funds. Kreander, Gray, Power and Sinclair (2005) studied the performance of 30 European SRI funds from four countries findings that there is no difference between SRI funds and conventional funds.

Bello (2005) studied 42 SRI U.S. mutual funds, he found no evidence of a performance difference between SRI and conventional funds. Both underperformed the Domini 400 Social Index and S&P 500 during the study period (1994 – 2001). Bauer, Koedijk and Otten (2005) investigated the performance of 32 British, 16 German and 55 U.S. SRI funds, they used Jensen and Carhart´s alpha and found that German and U.S. SRI mutual funds underperformed in both their relevant indexes and the conventional funds, whereas UK funds slightly outperformed, however the differences are not significant. Scholtens (2005) investigates the performance of Dutch SRI funds and finds that these funds outperformed conventional funds but with no statistically significant difference.

Barnett and Salomon (2006) studied 61 SRI funds tracked by the US Social Investment Forum (USSIF). They found that the relationship between financial and social performance is neither strictly negative, nor strictly positive. Instead, they found a curvilinear relationship, suggesting that two viewpoints may be complementary. Risk-adjusted performance varies with the types of social screens used. Community relations screening (excludes firms that do not invest in and/or develop economically depressed communities) increased financial performance, but environmental and labor relations screening (excludes firms with a record of poor environmental performance and firms with a record of poor labor relations practices, respectively) decreased financial performance.

Bauer, Otten and Rad (2006) investigated the performance of Australian ethical funds, and Bauer, Derwall and Otten (2007) evidence from Canada finding no statistical difference in performance between these two types of funds. Gregory and Whittaker (2007) in the UK market found that neither SRI nor non-SRI funds exhibited significant under performance on a risk/style adjusted under any of the models.

Renneboog, Ter Horst and Zhang (2008a) found that SRI funds in the US, the UK, and in many continental European and Asia-Pacific countries underperformed their domestic benchmarks. However, with the exception of France, Japan and Sweden, the risk-adjusted

performance of SRI funds is not statistically different from the performance of conventional funds. Cortez, Silva and Areal (2011) performed a study focused on 88 SRI funds from the European market from 1996 to 2007. They concluded that the performance of SRI funds is similar to the performance of both conventional and socially responsible indexes. Cortez, Silva and Areal (2012) Cortez Silva and Areal (2012) studied seven European markets and US market from 1996 to 2008. They found in several European markets (Belgium, France, Germany, Italy, the Netherlands and UK) that SRI funds showed similar performance compared to both conventional and benchmarks. In contrast, the US and Austrian funds showed evidence of underperformance. Nofsinger and Varma (2014) found that SRI mutual funds outperformed conventional funds in the global financial crisis, so they can be an optimal choice for investors who want to protect themselves from downside risk. They also found that SRI funds underperform at other times. Leite and Cortez (2014) performed a multi-country study focused on 54 international SRI funds located in eight European markets (Austria, Belgium, France, Germany, Italy, the Netherlands, UK and Spain), they applied the five-factor model and found a similar performance between socially responsible funds and conventional funds. Muñoz, Vargas and Marco (2014) studied 89 European green funds and 18 US funds, from 1994 to 2013. They applied the Carhart four-factor model and stated that for the US market, green funds did not perform any worse than the market, but with a global equity portfolio green funds showed evidence of underperformance. Becchetti, Ciciretti, Dalo and Herzel (2015) find no clear-cut dominance over the entire period (1992-2012) but also find that SRI funds generally did better than conventional funds in the period following the global financial crisis of 2007. Leite and Cortez (2015), focusing on the French market, found that SRI funds underperformed slightly more than their matched-samples according to different models, but differences in alphas are not statistically significant in most cases. They only found significance in one of the estimated models at the 10% level. Humphrey, Warren and Boon (2016) found that SRI managers have longer tenure and are more likely to be female, but they did not find any significant difference in the performance of SRI and conventional funds.

A recent and relevant reference is El Ghoul and Karoui (2017), to our knowledge this is the only paper that does not employ a dichotomous criterion in the selection of SRI mutual funds. Authors employ a corporate social responsibility score (CSR score), which is an asset-weighted composite CSR fund score. They showed the effects of CSR on fund performance, compared to low-CSR funds high-CSR funds displayed a poorer performance. In Table 23 we have summarized previous research on SRI fund performance.

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

Table 23.- Previous research on SRI fund performance

Authors	Sample (country and time spam)	Fund Sample	Relationship	Conclusions
Luther, Matatko and Corner (1992)	*UK; 1984-1990*	*15 SRI funds*	*Positive*	*-SRI funds outperform the index. There is clear evidence that the "ethical" trusts have UK investment portfolios more skewed towards companies with low market capitalization than the market as a whole.*
Hamilton, Jo and Statman (1993)	*US; 1981--1990*	*32 SRI funds*	*Neutral*	*-In general, the average of performance is similar between SRI funds and conventional funds.*
Luther and Matatko (1994)	*UK; 1985-1992*	*9 SRI Funds*	*Neutral*	*-No significant difference between SRI funds and conventional funds*
White (1995)	*US and Germany, 1991-1993*	*6 US funds and 5 German SRI funds*	*Negative*	*-SRI investments underperform the benchmark in different performance measures*
Mallin, Saadouni and Briston (1995)	*UK; 1986-1993*	*29 SRI Funds*	*Neutral Positive*	*-Ethical funds performed as well as their non-ethical counterparts and better than the non-ethical funds when the Jensen performance measure was used*
Gregory, Matatko and Luther (1997)	*UK, 1986-1994*	*18 SRI funds*	*Neutral*	*-No significant difference in financial performance exists between the two groups of funds*
Statman (2000)	*US; 1990-1998*	*31 SRI funds*	*Neutral*	*-Most of the SRI funds have a similar performance as the S&P 500-index and the DSI. SRI funds exhibit a positive but not significant Jensen´s alpha relative to the conventional funds*
Kreander, Gray, Power and Sinclair (2002)	*Belgium, Germany, Netherlands, Norway, Sweden, Switzerland, UK; 1986-1998*	*40 SRI Funds*	*Neutral*	*-On average, the SRI funds gave the same returns as conventional funds. There is some evidence that ethical funds are less risky as measured by volatility of returns and fund beta than their non-ethical counterparty.*
Kreander, Gray, Power and Sinclair (2005)	*Germany, Sweden, Netherlands and UK 1995-2001*	*30 SRI funds*	*Neutral*	*-There is no difference between ethical and non-ethical funds according to the performance measures employed*
Bello (2005)	*US; 1994 – 2001*	*42 SRI funds*	*Neutral*	*-No significant difference in investment performance between SRI and conventional funds.*
Bauer, Koedijk and Otten (2005)	*German, UK, US, 1990-2001*	*103 SRI funds*	*Neutral*	*-SRI mutual funds underperform the conventional funds; however, differences are not statistically significant*
Scholtens (2005)	*Netherlands, 2001-2003*	*12 SRI funds*	*Neutral*	*-SRI funds outperform conventional funds but with no statistically significant difference.*
Barnett and Salomon (2006)	*US, 1972-2000*	*61 SRI funds*	*Negative/ Positive*	*-Community relations screening increased financial performance. Environmental and labour relations screening decreased financial performance*
Bauer, Otten and Rad (2006)	*Australia, 1992-2003*	*25 SRI funds*	*Negative Neutral Positive*	*-During 1992-1996 domestic ethical funds underperformed conventional funds while international ethical funds outperformed conventional funds. During 1996-1999 domestic ethical funds outperformed the performance of conventional funds. No statistically significant difference between SRI funds and regular funds was found during 1999-2003 or in the whole period (1992-2003).*

Continue on next page

Table 23 (Cont.).- Previous research on SRI fund performance

Authors	Sample (country and time spam)	Fund Sample	Relationship	Conclusions
Bauer, Derwall and Otten (2007)	*Canada, 1994-2003*	*8 SRI funds*	*Neutral*	*-Ethical funds underperformed conventional funds, but the performance differential is statistically insignificant*
Gregory and Whittaker (2007)	*UK, 1989-2002*	*32 SRI funds*	*Neutral*	*-No significant difference in investment performance between -- SRI and conventional funds.*
Renneboog, Ter Horst and Zhang (2008a)	*23 countries and offshore jurisdiction (US, UK, continental European and Asia-Pacific countries), 1991 2003*	*463 SRI funds*	*Negative Neutral*	*-The average SRI fund in most European and Asia Pacific countries strongly underperform their benchmark portfolios. -In UK and US, the risk adjusted returns of SRI funds are not significantly different from conventional funds.*
Cortez, Silva and Areal (2011)	*Austria, Belgium, France, Germany, Italy, the Netherlands and UK, 1996 to 2007*	*88 SRI funds. Index benchmarks*	*Neutral*	*-Performance of SRI funds is similar to the performance of conventional benchmarks (MSCI AC World Index and MSCI AC Europe Index) and socially responsible benchmarks (FTSE4Good Global Index and FTSE4Good Europe Index)*
Cortez, Silva and Areal (2012)	*European market (Austria, Belgium, France, Germany, Italy, the Netherlands and UK) and US market, 1996-2008*	*39 funds for European markets and 7 funds for US market*	*Negative Neutral*	*-SRI funds for Belgium, France, Germany, Italy, the Netherlands and UK show similar performance compared to both conventional and benchmarks. -In contrast, for US and Austrian funds they showed evidence of underperformance*
Nofsinger and Varma (2014)	*US, 2000-2011*	*240 SRI funds*	*Neutral Positive*	*-Alphas for the SRI funds are not significantly different than conventional fund alphas. -It is a slight insignificant underperformance compared to conventional funds during the non-crisis periods. During the crisis periods (arch 2000 to October 2002 and October 2007 to March 2009), the SRI funds outperformed the conventional funds.*
Leite and Cortez (2014)	*Austria, Belgium, France, Germany, Italy, the Netherlands, UK and Spain, 2000 2008*	*54 SRI funds*	*Neutral*	*-Differences in the performance of international SRI funds and their conventional peers are not statistically significant*
Muñoz, Vargas y Marco (2014)	*US and European countries, 1994-2013*	*18 US and 89 European green funds*	*Negative Neutral*	*-For US SRI funds obtain statistically significant performance in crisis periods but underperform relative to the market in non-crisis periods. For European SRI funds obtain statistically insignificant performance irrespective of market conditions.*
Leite and Cortez (2015)	*France, 2000-2008*	*50 SRI funds*	*Negative Neutral*	*-SRI funds underperformed slightly more than their matched-samples according to different models, but differences in alphas are not statistically significant in most cases*
Becchetti, Ciciretti, Dalo and. Herzel, (2015)	Global,1992–2012	1,213 SRI funds	Positive	*-Socially responsible funds played an outperformed conventional funds during the 2007 global financial crisis*
El Ghoul and Karoui (2017))	US, 2003-2011	2,168 funds	Negative	*-The CSR score of the portfolio is negatively related to risk-adjusted performance.*

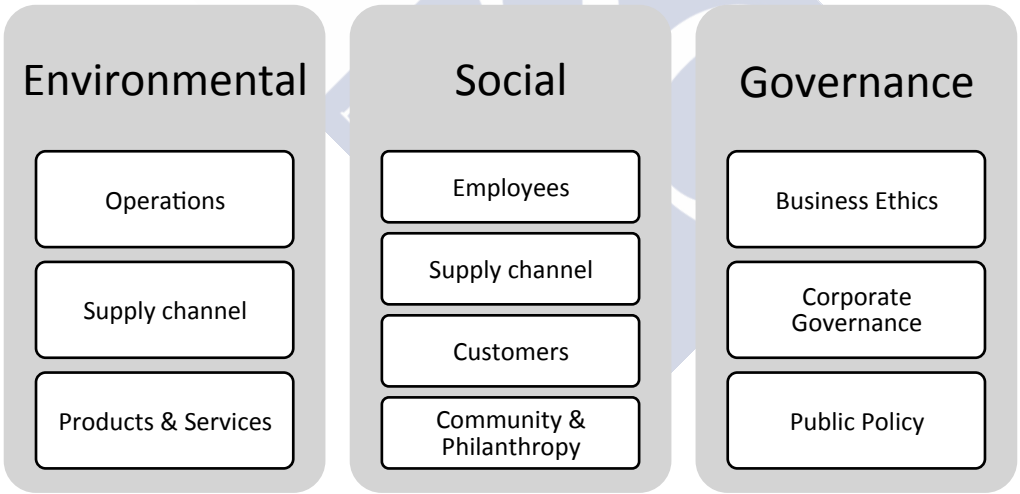
3. Background of Sustainalytics´ Methodology and Morningstar

3.1 Sustainability Scores.

3.1.1. Sustainalytics´ Methodology

Sustainalytics is a global company leader in ESG research and analysis. Sustainalytics´ ESG methodology consists of approximately 150 ESG indicators[7] to measure a company's sustainable practices. Sustainalytics assess company performance based on several internal and external data sources (Sustainalytics, 2016): review of company reporting (annual reports, etc.), review of external sources (NGOs, publications, etc.), analysis is done by an experienced analyst, structural peer review, company feedback and research process. The three pillars of the ESG Score are Environment, Social, and Governance. In each pillar several categories are distinguished (Figure1), for example in the Pillar Environment you can see Operations, Supply channel and Products and Services. Within these categories there are various indicators.

Figure 1- Sustainalytics Framework.



Source: Sustainalytics Framework.

Sustainalytics use two kinds of indicators templates: core and sector-specific. Core indicators are those used for all companies. Sector-specific indicators are those used to adjust to sector specific characteristics when it is not considered. There are two main types of scoring schemes for indicators: binary and linear. Binary indicators are those which the possible raw scores for binary indicators are 0 or 100. Linear indicators: are those which

[7] The exact number depends on the company´s industry.

there are various possible raw scores (0, 25, 50, 75 or 100). For the ESG indicators, Sustainalytics differentiate between three types that focus on different dimensions: preparedness, disclosure and performance (Sustainalytics, 2016):

- Preparedness indicators assess if company management systems and policies are well designed to manage material ESG risks.
- Disclosure indicators assess if a company reporting meets International best practice standards and if is transparent with respect to most material ESG issues.
- Performance indicators assess ESG performance based on quantitative metrics (for example carbon intensity) and qualitative based on the analysis of controversial incidents. There are special indicators that assess whether companies are involved in certain controversies. Controversies fall into five categories: category 1 – low, 2 – moderate, 3 – significant, 4 – high and 5 – severe. Controversy topics include: Business Ethics, Society and Community, Environmental Operations, Environmental Supply Chain, Product and Service, Employee, Social Supply Chain, Customer, Governance, and Public Policy.

Finally, to calculate the total score of the company, as well as aggregate scores on the three pillars, Sustainalytics uses a default weight matrix that is uniquely defined for every peer industry group (42 different comparable sub-industries).

3.1.2. Morningstar Sustainability and ESG Scores

The Morningstar Portfolio Sustainability Score is a measure developed in 2016 for scoring mutual funds and ETFs about environmental, social, and governance, or ESG, risks and opportunities. The subsequent Morningstar Sustainability Rating is a comparison relative to their Morningstar Category peers and is derived from the Morningstar Portfolio Sustainability Score. Morningstar Portfolio Sustainability Score (Sustscore) is defined as follows (Morningstar, 2016a and 2016b):

$$SustScore = Portfolio\ ESG\ Score - Portfolio\ Controversy\ Deduction$$

In order to receive a portfolio sustainability score, a portfolio must have a portfolio ESG score and a portfolio controversy score, which according to Morningstar (2016b) at least of 50% of a portfolio's assets under management must have a company ESG score and a controversy score/deduction. Based on their portfolio Sustainability score, ESG scores and

controversy scores, and descriptive ranks within their Morningstar Categories, funds are ranked as 5 (High), 4 (Above Average), 3 (Average), 2 (Below Average) and 1 (Low) (see Table 24 and 25).

Table 24- Portfolio Sustainability score and Sustainability Rating.

Distribution	Score	Descriptive Rank
Highest 10% (best)	5	High
Next 22.5%	4	Above Average
Next 35%	3	Average
Next 22.5%	2	Below Average
Lowest (worst)	1	Low

Source: Morningstar (2016b).

Table 25-Portfolio Controversy Scores

Distribution	Score	Descriptive Rank
Lowest 10% (Best)	5	High
Next 22.5%	4	Above Average
Next 35%	3	Average
Next 22.5%	2	Below Average
Highest 10% (Worst)	1	Low

Source: Morningstar (2016b).

Morningstar Portfolio ESG Score(ESGscore)[8] is calculated as:

$$\text{ESGscore} = \sum_{i=1}^{n} \text{w}_\text{i}\text{ESGNorm}_\text{i}$$

Where:

ESGNorm_i= the normalized ESG score of company i,

n= the number of securities in the portfolio,

w_i= the asset weight on security i, so the sum $\sum_{i=1}^{n} \text{w}_\text{i} = 100\%$.

To make the ESG scores comparable across peer groups, Morningstar normalizes the scores using a z-score transformation:

$$Z_i = \frac{\text{ESG}_\text{i} - \mu}{\delta}$$

Where:

ESGi = ESG score of company i,

μ = the mean of the ESG scores of the companies in the peer group,

δ = the standard deviation of the ESG scores of the companies in the peer group.

[8] Morningstar Portfolio Environmental Score, Social Score and Governance Score is calculated as an asset-weighted average of the scores of the individual companies.

Zi are used to create the normalized ESG scores[9] on a 0-100 scale, with a mean of 50, as:

$$\text{ESGNorm}_\text{i} = 50 + 10Zi$$

Normalized company ESG are aggregated to a portfolio ESG score using an asset-weighted average of all covered securities.

Sustainalytics tracks and categorizes ESG related incidents in companies, which is called "controversies". A single company may be involved in multiple ESG-related incidents at any given time, so Sustainalytics makes a controversy score of company i. Morningstar employs company controversy scores of Sustainalytics creating a Morningstar Portfolio Controversy Score (MContr_p), as follow:

$$\text{MContr}_p = \sum_{i=1}^{n} \text{w}_\text{i}\text{SCont}_\text{i}$$

Where:

w_i= the asset weight on security i

SCont_i = the Sustainalytics controversy score of company i.

4. Empirical study

4.1. Sample

Our sample contains 1,593 European equity funds rated by Morningstar in November 2016. The funds are the "open funds" type with ESG score in the investment area of Europe. Investment area identifies the geographic area where the fund focuses their investments. Furthermore, to avoid problems of multicollinearity, we have selected only an equivalent class for each fund. We obtained for each equity mutual fund three different measures of performance and other variables such as size, volatility, social conscious, expenses and age. We also use the Morningstar style-box to control the effect of the different categories which are included in the sample. The number of funds varied when we consider the costs where the sample reduces from 1,593 to 571 motivated for the lack of data in Morningstar Direct.

[9] Normalized company ESG scores can be interpreted as follows:
(70-100) Company scores at least two standard deviations above average in its peer group
(60-70) Company scores one standard deviation above average in its peer group
50 Company scores at peer group average
(30-40) Company scores one standard deviation below average in its peer group
(0-30) Company scores at least two standard deviations below average in its peer group

4.2. Variables construction

Our sustainable variables have been obtained from Morningstar Direct, with original source of individual companies from Sustainalytics. We will employ five variables: three are the pillars scores [Environment score variable (Envscore), Social score variable (Socscore) and Government score variable (Govscore)], the fourth is the ESG score of a portfolio (ESGscore), and finally, the Portfolio Sustainability Score (Sustscore) which is the ESG score minus Portfolio Controversy Score.

We have divided the funds into two groups based on whether ESG scores are below or above the median. Then, we estimated the means and their differences between both groups. Table 26 reports the results of the univariate analysis. As can be observed, the differences are very significant between the two groups for the different scores, with a difference of approximately five points in favour of the funds included in the high score group.

Table 27 compares the funds declared sustainable (Socially conscious) and those that result from dividing the sample according to a low ESGscore or a high ESGscore criteria. As you can see, there are big differences but in general, mutual funds that are declared sustainable are from an ESGscore point of view too. However, there are many funds that are not declared sustainable but they are based on the level of sustainability of the companies that integrate the portfolio. Thus, by using scores investors have at their disposal a large number of funds that are not declared sustainable but that their portfolio is comparable to that of sustainable funds.

Table 26- Sustainability and ESG scores for different groups

Variable	Low ESG score	High ESG score	t-statistic
Sustscore	52.63	58.58	-39.29***
ESGscore	57.44	63.99	-33.78***
EnvScore	56.66	62.36	-30.70***
Socscore	57.01	62.95	-34.81***
GovScore	55.28	60.41	-31.97***

This table reports the values of sustainability variables considered in the analysis obtained from Morningstar direct database. The funds are classified into low or high groups depending on whether their score is above or below the median. The t-statistic for difference of means is reported in the third column. Sustscore is the level of sustainability of the mutual fund measured by Morningstar. ESGscore is the ESG score of a fund. EnvScore, Socscore and GovScore are the mutual fund scores for the three dimensions (environment, social and corporate governance). *Significant at 10%; ** significant at 5% and * significant at 1%.**

Table 27- Sustainability and ESG scores for different groups

	Sustainabledummy		
Sociallyconcious	**0**	**1**	**Total**
0	671	790	**1,461**
1	19	113	**132**
Total	**690**	**903**	**1,593**

This table reports the number of mutual funds classified as sustainable using two different dummy variables. Sustainabledummy is based on low or high sustainable scores depending on whether their score is above or below the median. Sociallyconscious is for those mutual funds declared as socially conscious.

4.3. Performance variables

We considered different performance measures from Morningstar database. Given that we only have ESG data available for December 2016 we have analyzed the performance and risk effect using the performance and risk metrics for the last two years based on Wimmer (2012) who shows that ESG-scores persisted for two years motivated by the changes in the holdings of the SRI mutual funds. In particular we used the raw return, alpha and Sharpe ratios.

The differences in performance between the high and low-CSR funds are negative when considering raw returns, Sharpe ratios and two years alphas. That is, higher ESG scores showed a poorer performance. Our results are consistent with those achieved by El Ghoul and Karoui (2017) for US mutual funds.

Table 28- Performance metrics for different groups

Metric	Low ESG score	High ESG score	t-statistic
Alpha-2y	0.030	-0.270	1.61*
Alpha-1y	-0.070	-0.200	0.49
Sharpe-2y	0.047	-0.004	4.34***
Sharpe-1y	0.009	-0.100	5.04***
Return-2y	7.450	6.370	5.27***
Return-1y	3.680	1.060	6.14***

This table reports the values of the performance metrics indicated in the first column, Alpha and Sharpe are risk adjusted returns calculated for two and one years estimated at the end of 2016. Return is the raw measure of profitability. The data has been obtained from Morningstar Direct database. The funds are classified into low or high groups depending on whether their score is above or below the median. The t-statistic for difference of means is reported in the third column. *Significant at 10%; ** significant at 5% and * significant at 1%.**

4.4. Downside risk variables

We also assessed fund performance by considering downside risk. Tail risk is commonly taken by mutual funds and it has been shown to be useful in explaining fund performance (Kelly and Jiang, 2014). Specifically, we examined whether sustainable mutual funds are more or less exposed to tail risk by measuring mutual fund downside risk by using the value-at-risk (VaR). VaR measures the maximum loss that a fund I can obtain for a given time period and a given confidence level (1-p) as:

$$Pr(R_i \leq VarR_i)$$

which is the loss associated with the p-th percentile of the return distribution. It can be computed as $Var_i = F_i^{-1}(p)$, where F_i is the return distribution of the fund i.

Table 29 shows the difference of means for downside risk measured by the historical monthly value at risk at 99% of confidence level. The evidence for VaR reveals that high-score mutual funds display less tail risk but is only statistically significant for the two years measured.

Table 29- Value at Risk for different groups

	Low ESG score	High ESG score	t-statistic
VaR2y	7.97	7.61	4.30***
VaR1y	6.96	6.83	1.31

This table reports the VaR values obtained from Morningstar Direct database. The funds are classified into low or high groups depending on whether their score is above or below the median. The t-statistic for difference of means is reported in the third column. *Significant at 10%; ** significant at 5% and * significant at 1%.**

4.5. Flow of funds

We measure the flow of funds as:

$$Flow_i = \frac{TNA_{i,t} - TNA_{i,t-1}\,(1+R_{i,t})}{TNA_{i,t-1}}$$

Where $TNA_{i,t}$ and $TNA_{i,t-1}$ are the total net assets for fund i at the end of year t and $t-1$,

respectively, and $R_{i,t}$ is the return of fund i in year t.

Table 30 displays the difference of means for the flow of funds showing positive differences for higher scored mutual funds.

Table 30- Flows of funds for different groups

	Low ESG score	High ESG score	t-statistic
Flowasset_2015	-0.01	0.02	-2.15**
Flowasset_2016	-0.10	-0.07	-1.75*

This table reports the values of flow of funds obtained from Morningstar Direct database. The funds are classified into low or high groups depending on whether their score is above or below the median. The t-statistic for difference of means is reported in the third column. *Significant at 10%; ** significant at 5% and * significant at 1%.**

4.6. Descriptive statistics

Table 31 shows the different variables considered in our work. As can be seen, the variables related to the level of sustainability have an average level close to 60 points, and the difference between the minimum and maximum is around 25 points. On average, the funds have a negative Alpha despite yielding positive returns for the term of 1 and 2 years. The average flow has been negative and the percentage declared to be socially responsible is very small (8%). The size is very variable, the expense ratio is greater than 1% because the mutual funds included invest in equity, and in general the funds have a high average antiquity.

Table 31- Descriptive statistics

Variable	Obs	Mean	Std. Dev.	Min	Max
Sustscore	1,593	56.01	4.20	40.20	66.08
ESGscore	1,593	61.16	5.02	44.89	70.76
Envscore	1,593	59.89	4.63	45.78	68.48
Socscore	1,593	60.38	4.48	45.61	69.66
GovScore	1,593	58.19	4.07	44.59	65.98
Alpha-2y	1,766	-0.12	3.51	-23.57	51.67
Alpha-1y	1,766	-0.07	5.79	-46.98	103.11
Sharpe-2y	1,829	0.03	0.26	-1.49	1.33
Sharpe-1y	1,829	-0.04	0.44	-2.68	2.34
Return-2y	1,829	7.01	4.51	-14.10	52.66
Return-1y	1,829	2.42	8.47	-41.86	110.74
VaR-2y	1,829	7.69	1.64	2.15	18.77
VaR-1y	1,829	6.86	2.01	0.85	19.32
Size_2015	1,953	307.00	811	0.00	18,900.00
Size_2016	1,834	289.00	683	0.89	14,800.00
Flowasset_2015	1,856	-0.00	0.24	-0.98	0.62
Flujoasset_2016	1,757	-0.08	0.23	-0.99	0.62
ExpRatio	765	1.34	0.73	0.00	5.21
Age	2,08	13.02	7.10	1.50	59.93
Sociallyconcious	2,08	0.08	0.27	0	1

This table reports the values of the variables considered in the analysis obtained from Morningstar direct database. Sustscore is the level of sustainability of the mutual fund measured by Morningstar. ESGscore is the ESG score of a fund. Envscore, Socscore and GovScore are the mutual fund scores for the three dimensions (Environment, Social and Corporate Governance). Alpha and Sharpe are risk adjusted metrics. Return is the raw profitability and VaR a downside risk measure. Size is the amount invested in millions of euros. Flowasset is the percentage of new funds

over total assets and ExpRatio the net expense ratio. The age are the years from inception date and Sociallyconcious is a dummy that takes the value 1 when the fund declares itself socially responsible and zero otherwise.

4.7. Fund Performance and Sustainability Scores

In this part, we test if the degree of sustainability measured through ESG scores has a positive or negative effect on performance. In addition, we consider environmental, social, and governance scores to evaluate the contribution of each dimension to the portfolio performance. We propose the following model:

$$Y_i = cons_i + \beta_1 Sustscore_i + \beta_2 Age_i + \beta_3 LossDev_i + \beta_4 logSIZE_i$$
$$+\beta_5 ExpRat_i + \beta_6 Sociallyconcious_i + \sum_j Category_j + \varepsilon_i$$

where:

Y_i= Alternative performance metrics for fund i.

i= 1 through N, where N is the total number of funds in the sample.

$Sustscore_i$ is the sustainability score provided by Morningstar.

Age = Years since inception date.

LossDev = standard deviation of mutual funds returns.

LogSize = logarithm of mutual fund market value.

ExpRat = Net expense ratio of fund i.

Sociallyconcious = dummy of SRI mutual funds.

Category= dummies of categories except small style

$cons_i$ and β_1,β_2,β_3,and β_4 are parameters of the regression and ε_i the term error.

Our results show that Sustscore is significant in explaining the level of performance for all the metrics and terms. If we use ESG scores instead of Sustscore the results are mainly the same. Most of the models present a negative sign in line with El Ghoul and Karoui (2017) and Renneboog *et al.* (2008a), who suggest that socially responsible mutual funds underperform other funds. The dummy variable is also significant, but to the contrary showing that considering the level of sustainability can help to understand better the relation between performance and social responsibility. Our results support Statman and Glushkov (2016) who conclude that the lack of clearly defined criteria to distinguish mutual funds as

socially responsible affects the results of previous research based on dichotomy variables. Among the control variables, Table 32 shows that volatility in all the models and the expense ratio, only in some of them, are negatively related to performance, while size and age are not significant.

Table 32- Sustainability Scores and fund performance

Variable	Return-2y	Return-1y	Sharpe-2y	Sharpe-1y	Alpha-2y	Alpha-1y
Sustscore	-0.2806***	-0.7726***	-0.0133***	-0.0351***	-0.0969***	0.0995*
logSize	0.1083	-0.1633	0.007	-0.0078	0.1570*	0.0169
Sociallyconcious	1.8588***	2.0932**	0.1110**	0.1055*	0.713	-0.4459
Age	0.0073	0.0499	0.0003	0.0027	-0.0029	0.0363
LossDev	-0.7611***	-1.3520***	-0.0335***	-0.0668***	-0.5550***	-0.6352***
ExpRatio	-0.4003	-0.4921	-0.0265*	-0.0374*	-0.4186	-1.4154***
Largeblend	-8.4605***	-12.5254**	-0.4801***	-0.6001**	-1.4166	-11.8063**
Largegrowth	-7.4792***	-14.8297**	-0.4052***	-0.7483***	-1.6033*	-13.3557***
Largevalue	-6.7732***	-6.835	-0.3787***	-0.2632	-0.3389	-9.6283*
Midblend	-6.0349***	-10.8419*	-0.3143***	-0.4648**	-1.3284	-10.4584**
Midgrowth	-4.5252***	-14.2118**	-0.2129***	-0.6378***	-1.6275*	-11.4390**
Midvalue	-9.9228***	-9.611	-0.5346***	-0.3633	-2.4733*	-10.5663**
cons	36.2064***	73.0547***	1.4123***	3.2451***	10.2035***	13.1380**
N	571	570	571	570	541	541
r2	0.3552	0.5621	0.3131	0.4951	0.1218	0.2219

This table reports the coefficients for the regression models for different performance measures. Alpha is the beta-adjusted return over a one-year period; Sharpe is the yearly risk-adjusted return and, Return is the total net return. Sustscore is the level of sustainability of the fund provided by Morningstar and Sociallyconcious is the common dummy variable used to analyse sociallyconscious mutual funds. N is the number of observations and r2 the R-squared fit measure. The dummies of categories have been included and compared with small mutual fund styles. *Significant at 10%; ** significant at 5% and * significant at 1%.**

Now using the different elements in which ESG scores are subdivided we have achieved similar results, finding in most models a negative relation between the dimensions of sustainability and performance. Again, those mutual funds with higher environmental scores reduce the level of performance adjusted and non-adjusted in five of the six models estimated. For the other dimensions (social and governance) the results are quite similar concluding that in general the effects of the different dimensions have a negative impact on alternative performance metrics.

Table 33- ESG Pillars models and fund performance.

Variable	Return2y	Return1y	Sharpe2y	Sharpe1y	Alphacat2y	Alphacat1y
Environment Score Models						
Envscore	-0.3373***	-0.7500***	-0.0169***	-0.0343***	-0.0747**	0.1474**
logSize	0.1717**	-0.1821	0.0114**	-0.0086	0.2491***	0.1248
Sociallyconcious	1.6921**	1.1247	0.1058**	0.0621	0.5883	-0.4186
Age	0.0082	0.0492	0.0004	0.0027	-0.0026	0.0363
LossDev	-0.6834***	-1.3802***	-0.0296***	-0.0680***	-0.5195***	-0.6579***
ExpRatio	-0.3602	-0.4927	-0.0235	-0.0376*	-0.3068	-1.2899***
_cons	37.2674***	72.6572***	1.4858***	3.2394***	6.8766**	8.683
N	571	570	571	570	541	540
r2	0.3791	0.5443	0.3379	0.4834	0.1239	0.2283
Social Score Models						
Socscore	-0.3071***	-0.7310***	-0.0151***	-0.0335***	-0.0942***	0.1341**
logSize	0.1559*	-0.2072	0.0106*	-0.0098	0.2445***	0.1271
Sociallyconcious	1.7537**	1.4089	0.1081**	0.0751	0.6708	-0.4506
Age	0.0094	0.053	0.0005	0.0029	-0.0015	0.0356
LossDev	-0.7240***	-1.4174***	-0.0316***	-0.0697***	-0.5324***	-0.6445***
ExpRatio	-0.3423	-0.4659	-0.0225	-0.0363	-0.3108	-1.3014***
_cons	36.9893***	73.9977***	1.4561***	3.3005***	8.3183***	8.935
N	571	570	571	570	541	540
r2	0.3644	0.5429	0.3252	0.4823	0.1291	0.227
Government Score Models						
Socscore	-0.1133***	0.1867***	0.0012	-0.0318***	0.0012	-0.0318***
logSize	0.1907***	0.02	0.0139***	-0.011	0.0139***	-0.011
Sociallyconcious	0.0874	-0.1171	0.0507	0.0665	0.0507	0.0665
Age	-0.0244**	0.0003	-0.0007	0.0033	-0.0007	0.0033
LossDev	-0.5027***	-0.7989***	-0.0382***	-0.0712***	-0.0382***	-0.0712***
ExpRatio	-	-	-0.0248*	-0.0423*	-0.0248*	-0.0423*
_cons	10.3321***	7.5867	0.1834	3.1898***	0.1834	3.1898***
N	723	729	571	570	571	570
r2	0.1128	0.2555	0.3358	0.4698	0.3358	0.4698

This table reports the coefficients for the regression models for different performance measures. Alpha is the beta-adjusted return over a one-year period; Sharpe is the yearly risk-adjusted return and, Return is the total net return. Sociallyconcious is a dummy variable used to analyse sociallyconscious mutual funds. N is the number of observations and r2 the R-squared fit measure. *Significant at 10%; ** significant at 5% and * significant at 1%.**

4.8. Downside risk and sustainability scores

In this part, we test if the degree of sustainability measured through ESG scores and their components, has a positive or negative effect on the historical value at risk of the portfolio.

We used the following model:

$$VaR_i = cons_i + \beta_1 Sustscore_i + \beta_2 Age_i + \beta_3 LossDev_i + \beta_4 logSIZE_i +$$

$$\beta_5 ExpRat_i + \beta_6 SocRespdum + \sum_j Category_j + \varepsilon_i$$

As Table 34 shows, the downside risk of mutual funds is affected by the level of sustainability (ESG score). Specifically, we observed how the variable Sustscore is

negatively and significantly related with the VaR of the fund at 99% of confidence level in both terms, one and two years. These results support that funds with a higher degree of sustainability protect investors better against extreme losses. As Kurtz (1997) or Goldreyer and Diltz (1999) explain, SRI mutual funds managers base their decisions on a deeper, more complete and higher quality information, resulting in a significant reduction in the risk of their investment decisions. On the other hand, the dichotomous variable commonly used has a positive and opposite sign to that resulting from using a continuous variable. We also made the analysis for the different subfactors, observing again a negative and significant relationship for most of the estimated models. As can be seen in Table 12, the increase in the level of environmental, social and governance sustainability reduces the level of extreme losses of investment funds. It is again observed that the dummy variable is significant and positively related to the level of risk. From this analysis, we observed that the results of evaluating the effect of sustainability based on dichotomous variables may yield contradictory results to those obtained when continuous variables are used.

Table 34- Sustainable score and downside risk

Variable	VaR-2y	VaR-1y	VaR-2y	VaR-1y	VaR-2y	VaR-1y	VaR-2y	VaR-1y
Sustscore e	-0.0211*	-0.0277***	-	-	-	-	-	-
EnvScore	-	-	-0.0193	-0.013	-	-	-	-
Socscore	-	-	-	-	-0.0256**	-0.016		
GovScore	-	-	-	-	-	-	-0.0289**	-0.0148
logSize	-0.0231	-0.0278	-0.0236	-0.0278	-0.025	-0.0285	-0.0258	-0.0291
Sociallyconcious	0.1671*	0.3396**	0.1397	0.2730*	0.1637*	0.2872*	0.1698*	0.2820*
Age	0.0083**	0.0149***	0.0082**	0.0148***	0.0085**	0.0149***	0.0089**	0.0151***
LossDev	0.7757***	0.6817***	0.7805***	0.6745***	0.7777***	0.6753***	0.7798***	0.6744***
ExpRatio	0.0185	0.0472	0.018	0.0536	0.0163	0.0526	0.0094	0.05
Largeblend	-1.1074***	-0.5057	-1.0197***	-0.5082	-1.0188***	-0.5148	-0.9981***	-0.5246
Largegrowth	-1.0165***	-0.6747*	-0.9683***	-0.7100*	-0.9533***	-0.7043*	-0.9541***	-0.7232*
Largevalue	-1.2509***	-0.5573*	-1.1523***	-0.5113	-1.1715***	-0.5302	-1.1336***	-0.5225
Midblend	-0.7316***	-0.538	-0.7191***	-0.5515	-0.7328***	-0.5607	-0.7162***	-0.5559
Midgrowth	-0.6022***	-0.7343**	-0.6060**	-0.7415*	-0.6009***	-0.7383**	-0.5998***	-0.7388**
Midvalue	-1.0851***	-0.7785**	-1.0145***	-0.6938	-1.0971***	-0.7441*	-1.0475***	-0.7105*
cons	2.4437***	2.7915***	2.3131**	2.0917***	2.7512***	2.2913***	2.8750***	2.2177***
N	571	570	571	570	571	570	571	570
r2	0.72	0.85	0.72	0.85	0.72	0.85	0.72	0.85

This table reports the coefficients for the regression models. VaR is the maximum loss that a fund i can obtain for a given time period and a given confidence level. Sustscore is the level of sustainability of the mutual fund measured by Morningstar. ESGscore is the ESG score of a fund. EnvScore, Socscore and GovScore are the mutual fund scores for the three dimensions (environment, social and corporate governance). Sociallyconcious is a dummy variable used to analyse Sociallyconcious mutual funds. N is the number of observations and r2 the R-squared fit measure. *Significant at 10%; ** significant at 5% and * significant at 1%**

4.9. Flows and sustainability scores

In this section, we analyze the effect of sustainability on the flows of investment funds. In particular, flows of sustainable funds are generally considered to be less sensitive to changes in performance because investors value other elements in their utility function. Benson and Humphrey (2008) and Renneboog *et al.* (2011) obtain evidence in favor of greater stability in flows for sustainable funds, while Bollen (2007) finds that SRI mutual funds are more sensitive to positive returns and less to negative ones. In line with El Ghoul and Karoui (2017) we argue that funds with higher ESG scores attract more conscious investors less worried about performance and therefore the flows are less sensitive to past performance. Thus, we estimate the following model to evaluate the effect of sustainability on the flow of funds using the different performance metrics (alpha, Sharpe, return), the sustainability score and the interaction of the product (SustPerf: sustsharpe, sustalpha or sustreturn):

$$Flow_i = \alpha_i + \beta_1 Performance_i + \beta_2 Sustscore_i + \beta_3 SustPerf_i + \beta_4 Age_i + \beta_5 LossDev_i$$

$$+\beta_6 logSIZE_i + \beta_7 ExpRatio_i + \beta_8 \text{Sociallyconcious} + \sum_j Category_j + \varepsilon_i$$

Where:

SustPerf: is the product of Sustscore and sharpe (sustsharpe), alpha (sustalpha) or net return (sustreturn) depending on the model.

Table 35 shows that only the model that takes as performance variable the profitability without risk adjustment, is significant. This would indicate that unadjusted returns are the ones that have had the most influence on investment decisions. On the other hand, in model three, the sustainability score is also significant, so that higher-rated funds received a larger volume of funds than those with a lower score. This fact shows that the degree of sustainability stimulates fund raising and more when the degree of sustainability is higher. Also, when we analyze the effect of the sustainability dummy variable (Sociallyconcious) it is significant in all models, which confirms the importance of sustainability in attracting investors interested in funds that are declared sustainable. This fact can be related to both greater social awareness and expectations of greater profitability in socially responsible investments. Finally, the negative sign of the interaction variable (sustreturn) shows the lower sensitivity of sustainable funds, supporting the results found by El Ghoul and Karoui (2017) using alternative metrics and US funds.

Table 35- Sustainable score and flow of funds

Variable	Model1	Model2	Model3
Sharpe-2y	0.3936	-	-
Alpha-2y	-	0.0143	-
Return-2y	-	-	0.0224***
sustsharpe	-0.0029	-	-
sustalpha	-	-0.0001	-
susreturn	-	-	-0.0002**
Sustscore	0.0028	0.002	0.0090***
Sociallyconcious	0.1014***	0.1129***	0.1045***
logSize	0.0243**	0.0221**	0.0295***
Age	0.0001	0.0006	-0.0004
LossDev	-0.0148**	-0.0158***	-0.0056
ExpRatio	0.04	0.0382	0.0438*
Largeblend	-0.2039**	-0.3165***	-0.1727
Largegrowth	-0.2269**	-0.3007***	-0.1537
Largevalue	-0.1957**	-0.2704***	-0.1977
Midblend	-0.2237**	-0.2511***	-0.1624
Midgrowth	-0.2316***	-0.2645***	-0.1222
Midvalue	-0.1913	-0.2839***	-0.1823
cons	-0.4622*	-0.2827	-1.0624***
N	560	531	560
r2	0.084	0.0709	0.0995

This table reports the coefficients for the regression models. Alpha is the beta-adjusted return over a two-years period; Sharpe is the yearly risk-adjusted return and, Return is the total net return. Sustscore is the level of sustainability of the mutual fund measured by Morningstar. Sociallyconcious is a dummy variable used to analyse socially conscious mutual funds. N is the number of observations and r2 the R-squared fit measure. *Significant at 10%; ** significant at 5% and * significant at 1%.**

5. Robustness

We conducted some additional robustness tests to check the consistency of our results and to provide other complementary analysis. We checked whether performance may differ attending to the fund manager skills considering the quantiles of different performance measures: differences in the quantiles would indicate differences in the fund manager's abilities to deal with performance.

Quantile regression let capture information about the coefficients at different quantiles of the dependent variable given the set of endogenous variables. In addition, the conditional quantile regression developed by Koenker and Bassett (1978) deals well with skewed distributions of fund performance. In particular, we adopted the bootstrapping method proposed by Efron (1979) and implemented in the software Stata 12. Given Y_i as the different performance metrics used in this paper, and X_i as a vector of exogenous variables representing the sustainable score of each mutual funds and other controls, the quantile model can be written as:

$$y_i = X_i'\beta_\phi + u_{\phi i}$$

Assuming that:

$$Quant_\phi(y_i|X_i) = X_i'\beta_\phi$$

$$Quant_\phi(u_{\phi i}|X_i)=0$$

Table 36 reports quantile parameter estimates for three different adjusted risk-return performances. Our evidence for all quantiles confirms no differences in the results and sustainability seems to be important independent of the level of performance analysed.

We also calculated the models excluding the expense ratio because this variable has many blanks and reduces the sample a lot. After the calculations, we again observed no differences with the models presented in the previous empirical analysis. Finally, we recalculated the models for each category and we obtained different results depending on the category, concluding that on average the effect is negative on performance but specific for each category.

Table 36- Quantile regression.

	Return2y	Return1y	Sharpe2y	Sharpe1y	Alphacat2y	Alphacat1y
q25						
Sustscore	-0.1842***	-0.3622***	-0.0115***	-0.0205***	-0,0046	0.1326**
logSize	0.1722*	-0,1016	0,0086	-0,01	0.2409**	-0,0609
Sociallyconcious	0,999	0,4466	0,06	0,0575	0,2995	0,0173
Age	0,0149	0,0525	0,0008	0.0039*	0,0044	0.0675*
LossDev	-0.7367***	-1.1443***	-0.0234**	-0.0511***	-0.4590***	-0.6372***
ExpRatio	-0.8319***	-0.9671***	-0.0515**	-0.0706***	-1.1022***	-1.7055***
cons	30.6208***	39.3289***	1.0995***	2.0197***	3,7736	4,2163
q50						
Sustscore	-0.2361***	-0.6639***	-0.0131***	-0.0356***	-0.0687***	0,077
logSize	0.1673**	-0,0868	0,0089	-0,0079	0.2665***	-0,032
Sociallyconcious	1,0751	1.8932*	0,0815	0.1372**	0,7089	-0,1515
Age	0,0103	0,0626	0,0007	0,0043	-0,0109	0,0347
LossDev	-0.4361***	-1.2064***	-0.0204***	-0.0631***	-0.3271***	-0.6350***
ExpRatio	-0.6436**	-0.7434*	-0.0364**	-0.0661***	-0.4729**	-1.6412***
cons	30.4631***	57.1273***	1.3114***	3.0006***	4,6984	11,0889
q75						
Sustscore	-0.3204***	-0.8217***	-0.0157***	-0.0365***	-0.0992***	-0,0348
logSize	0,0408	-0,0436	-0,0024	0,0105	0,0465	-0,0804
Sociallyconcious	1.7210**	2,3816	0.1026*	0,0727	0,9631	-0,0929
Age	-0,0272	0.0826*	-0,0014	0,0042	-0,0083	0,0065
LossDev	-0.6449***	-1.5050***	-0.0331***	-0.0803***	-0.3354***	-0.4743***
ExpRatio	-0,442	0,3107	-0,0344	0,0134	-0.5127*	-1.1507***
cons	39.7659***	83.9232***	1.7920***	3.6249***	10.7921***	27.0078**
N	571	570	571	570	541	541

This table reports the coefficients for the quantile regression models. Sustscore is the level of sustainability of the mutual fund measured by Morningstar. Sociallyconcious is a dummy variable used to analyse socially conscious mutual funds. N is the number of observations. *Significant at 10%; ** significant at 5% and * significant at 1%.**

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

6. Conclusion

Socially Responsible Investment (SRI) summaries any investment strategy which search for a financial return and encourage corporate practices that promote environmental care, consumer protection and human rights. In Europe, SRI strategies grew by 11.7% from 2014 to 2016 to reach $12.04 trillion. Traditional studies focus their work on mutual funds which declare themselves as funds that support a SRI approach. One important limitation of this approach is that results could be biased, because SRI mutual funds could have different levels of sustainability and differences with conventional funds could be not significant. Recently, Morningstar launched Morningstar Sustainability Score to classifying mutual funds about ESG factors. The use of sustainability scores in our work can allow us to evaluate the effect of the degree of sustainability on performance, risk or flows on European equity mutual funds.

Our result shows that there are a large number of funds that are not declared sustainable but their portfolio is comparable to sustainable mutual funds. Furthermore, Sustainability score is significant explaining the level of performance for all the metrics analysed (alpha, Sharpe and net return), with negative sign in most models. Using a conventional dummy to declare social mutual funds, the results are significant but with the contrary sign, showing that considering the level of sustainability can help to understand better the link between performance and social responsibility. Our results are in accordance to Statman and Glushkov (2016), who concluded that the lack of clearly defined criteria to distinguish SRI mutual funds affected the results. Also, we obtained similar results to El Ghoul and Karoui (2017) for US mutual funds market. Using the different pillars of ESG scores (environmental, social, and governance) we were able to achieve a negative link between the dimensions of sustainability and performance, showing that all the dimensions play an important role in explaining performance.

In terms of downside risk, the level of sustainability is negatively and significantly related to the VaR of the fund, supporting that higher scored mutual funds protect better against extreme losses. The opposite is found for the conventional dummy, showing the advantages of employing a quantitative measure of sustainability to evaluate assets´ risk. This result could mean that SRI mutual funds managers base their decisions on a deeper analyse resulting in a significant reduction in the risk of their investment decisions. Our work shows that sustainability scores can be used by investors worried by extreme losses and not only by values-motivated investors.

Finally, we analyzed the effect of sustainability on the flows realizing that unadjusted returns have the most influence on investment decisions. The sustainability score is significant on the flows, so higher-rated funds received a larger volume of funds and it is also significant the

effect of the sustainability dummy variable. Finally, the negative sign of the interaction variable (product of sustainability and return) shows the lower sensitivity of sustainable funds. This shows the different sensitivity to performance of values-motivated investors.

The limitations of our work are that we do not have a panel database, so we only have the observed values of sustainability and ESG scores on a data point (December 2016). Future research could use panel data and mix the Morningstar ESG scores with the MSCI ESG scores to analyze the effect of several scores.

Chapter V. Conclusions, limitations and future research

Mutual funds are a product highly requested by investors and savers. Our work seeks to find aspects that can predict the selection of the best funds, more specifically, we intend to look at the extent to which quantitative, qualitative and social/environmental ratings can help to disentangle the funds that will have a better performance in the future, as well as preserve long-term wealth. Choosing not only the funds that get better risk adjusted returns, but also those that are more robust in adverse times.

In this study, the Chapter II focuses on quantitative ratings (Star Ratings), the Chapter III focuses on the qualitative ratings (Analyst Ratings) and finally, the Chapter IV focuses on socially responsible scores. All chapters, are based in Morningstar ratings because it is the most important mutual fund research firm and of important matter to the investors, savers, financial institutions and academics, for its independence and specialization in mutual funds.

Morningstar quantitative ratings, which evaluate funds with stars as do hotels (1 to 5), has become a very popular tool for investors. Qualitative and Sustainability ratings are not yet as widely used by the fund industry as they are relatively recent, especially the latter.

Many investors select their investments in mutual funds based exclusively on the quantitative rating. In this sense, we wanted to verify if the exclusive use of this criterion that allows making good decisions related to performance and downside risk. We selected European equity funds, we used Rating Overall and also three years ratings to check the robustness of our estimates.

Our conclusions supports the ability of quantitative ratings to select funds that will behave better in terms of future performance. The results are in line with some previous empirical evidence found in Morey and Gottesman (2006), Müller and Weber (2014), and Meinhardt (2014). Consequently, we have found that on average, funds with a lower rating have a worse performance in terms of risk adjusted measures and Annual Return for the following year. The strongest predictability is observed one year ahead but it is also good for three-years. The inclusion of other variables such as costs, size and age reflects the importance of considering other variables for funds selection. Nevertheless, the ratings are still significant in explaining performance, indicating that costs are not the only factor that determines the predictive power of quantitative ratings. Finally, our analysis in terms of VaR (value at risk) shows that the investment in good rated funds can help to preserve the investors wealth better.

Our results also support the use of ratings in the investment funds selection process, accompanied by other quantitative variables. On the other hand, the greater significance achieved in the short term advises the review of portfolios on an annual basis.

Investment decisions on mutual funds based on quantitative rating, such as Star Ratings, can be improved by the inclusion of qualitative factors. Morningstar has two systems to classify mutual funds based on the qualitative aspects: Morningstar Analyst Rating and Morningstar Stewardship Grade. Its expectations are the prospects for quality, performance, quality, performance, quality of management, among others. Analyst Ratings is not so popular and very little research has been done. Kamal (2013) and Armstrong, *et al.* (2016) are the sole authors who focus their research on analyst ratings.

Our research attempted to assess to what extent the selection of mutual funds based on the Morningstar Analyst and Star Ratings criteria has an impact on investor's performance. We tried to understand to what extent having good analyst qualifications outperform those who do not recommended in the short (12 months) and long term (36 months), and if it is useful to combine both qualifications in the selection process to identify future bests mutual funds.

We found that there is little evidence that funds with a better Analyst Rating (Gold) have a better performance in terms of risk-adjusted measures (alpha and Sharpe). The results show predictability for one year ahead but not for three-years. This evidence is more relevant in the case of the analysis made by investment style´s category. In the analysis of the pillars in which the analyst ratings are broken down, we do not find evidence that future performance is related to any of these specific dimensions. The combination of Stars and Analyst ratings does have medium-term differentiation results, with a higher performance in terms of Sharpe's ratio for bronze (3, 4 or 5 stars) or three-star gold funds. The inclusion of other variables such as costs, size and manager tenure reflects the importance of considering, as it is in the case of the quantitative ratings, other variables for fund´s selection. Nevertheless, in several estimations Gold ratings are still significant in explaining performance, indicating that costs are not the only factor that determines the predictive power of qualitative ratings. Our results support the use of qualitative ratings in the investment funds selection process, accompanied by other variables.

Socially Responsible Investment (SRI) summarizes any investment strategy which search for a financial return and encourages corporate practices that promote environmental, consumer protection and human rights policies. Traditional studies focus their work on mutual funds which declare themselves as funds that support a SRI approach. One important limitation of this

approach is that results could be biased because SRI mutual funds could have different levels of sustainability and the differences with conventional funds could not be as significant.

The use of the Morningstar Sustainability Score to classifying mutual funds about ESG factors allowed us to evaluate the effect of the degree of sustainability on performance, risk or flows on European equity mutual funds. Our result shows that there are many funds that are not declared sustainable but their portfolio is comparable to sustainable mutual funds. Furthermore, Sustainability score is significant explaining the level of performance for all the metrics analyzed (alpha, Sharpe and net return), with negative sign in most models.

Using a conventional dummy to declare social mutual funds, the results are significant, but with the contrary sign, showing that considering the level of sustainability can help to better understand the relation between performance and social responsibility. Our results are in accordance to Statman and Glushkov (2016), who concludes that the lack of clearly defined criteria to distinguish SRI mutual funds affects the results. We have obtained similar results to El Ghoul and Karoui (2017) for US mutual funds market. Using the different pillars of ESG scores (environmental, social, and governance) we also achieve a negative relation between the dimensions of sustainability and performance, showing that all the dimensions play an important role in explaining performance. This phenomenon is known in the literature as the price to invest ethically.

In terms of downside risk, the level of sustainability is negatively and significantly related with the VaR of the fund, supporting that higher scored mutual funds protect better against extreme losses. The opposite sign is found for conventional dummy, showed the advantages of employing a quantitative measure of sustainability to evaluate the asset's risk. This result could mean that SRI mutual funds managers base their decisions in a deeper analysis resulting in a significant reduction in the risk of their investment decisions. Our work evidence that sustainability scores can be used by investors worried by extreme losses and not only by values-motivated investors.

Finally, we analyzed the effect of sustainability on the flows obtaining that unadjusted returns have the most influence on investment decisions. Sustainability score is significant on the flows, so higher-rated funds received larger volume of funds and it is significant the effect of the sustainability dummy variable. Finally, the negative sign of the interaction variable (product of sustainability and return) shows the lower sensitivity of sustainable funds which emphasizes the different sensitivity to performance of values-motivated investors.

Limitations and future research

This study has several inherent limitations because the use of data of mutual funds. Of course, there are many categories of funds, and for each category, the fund may have different currency or location, which makes it very difficult to control many variables. Besides, in the same category a fund can have very diverse portfolios.

Every fund has different classes, being able to have different institutional and non-institutional. Throughout the work we have tried to control these problems, choosing the same currency for all funds, the best class of each fund in terms of costs and taking into account in the category of the fund's regression.

Another limitation is the measurement of ratings. We know that the ratings are only rated with a minimum of 3 years in the case of Stars and many funds have a survived lower than that. This does not allow us to know the true reality of persistence in all funds, since we are excluding the youngest. That's why, in order to draw more objective conclusions, we use the overall rating for the Star Ratings quantitative ratings that focuses on the performance of the last 10 years. Although it allows us to make the analysis more robust, we are also exclude the most important parts of the funds from the study; the younger funds and the ones that did not survive over the years for being worse than the majority. The other limitations are due to the samples being that there are no long samples and the funds with Rating Analyst, are fewer when we consider the amount of funds that exist. The attribution of this Rating seems to be biased since there is practically no attribution of the Rating Negative, which is justified by Morningstar that this Rating is attributed to the best funds. Also, the sustainability ratings and scores are very recent, so they do not allow us to draw solid conclusions about them.

Mutual funds is a very addressed subject in literature with several measures of performance. Many of the measures are adjusted to the risk, as well as without adjustment. Risk-adjusted measures that are being used very frequently, in almost all investment fund studies that focus on performance metrics and are compared to the benchmark; we realize that this brings us problems of measurement. Many times, we are comparing portfolios that have nothing in common other than the objective of hitting the same benchmark, even when portfolio has nothing to do with the Index. That is why in this study we use the metrics considering the benchmark of the prospect and also of the fund category.

We also want to mention some other limitations of our work. In Chapter, we do not have a panel database, so we have only the observed values of sustainability and ESG scores on a data point (December 2016).

Future research can use panel data and mix the Morningstar ESG Scores with the MSCI ESG Scores to analyze the effect of several scores. There are also several studies to be carried out in the near future on investment funds and ratings. It will be interesting to look if the Stewardship ratings can help in the selection of funds. Also to see if sustainability ratings can help funds that have good grades to have better, worse, or neutral performance in relation to each other.

This study focuses on equity funds but it will also be interesting to verify if its results in other categories of funds such as ETFs, Index Funds or bound mutual funds. Among other things it would be interesting to use other performance metrics (alfas with three, four and five factors mainly).

References

Adams, J. C., Mansi, S. A. & Nishikawa, T. (2010). Internal governance mechanisms and operational performance: Evidence from index mutual funds, *Review of Financial Studies*, 23(3), 1261-1286.

Antypas, A., Caporale, G., Kourogenis, N., & Pittis, N. (2009). Selectivity, Market Timing and the Morningstar Star-Rating System. CESifo Working Paper Series No. 2580. Available at SSRN: http://ssrn.com/abstract=1360637.

Armstrong, W., Genc, E., & Verbeek, M. (2016). Going for Gold: An Analysis of Morningstar Analyst. Available at SSRN: https://ssrn.com/abstract=2419669

Baltagi, B. (2010). *Econometric Analysis of Panel Data*. 4th ed. London: JohnWiley & Sons.

Barnett, M.L. & Salomon, R.M. (2006). Beyond dichotomy: The curvilinear relationship between social responsibility and financial performance. *Strategic Management Journal*, 27(11), 1101–1122.

Bauer, R., Otten, R. & Rad, A.T. (2006). Ethical investing in Australia: Is there a financial penalty?, *Pacific-Basin Finance Journal*, 14(1), 33-48.

Bauer, R., Derwall, J. & Otten, R. (2007). The ethical mutual fund performance debate: new evidence from Canada. *Journal of Business Ethics*, 70(2), 111–124.

Bauer, R., Koedijk K. & Otten, R. (2005). International Evidence on Ethical Mutual Fund Performance and Investment Style. *Journal of Banking and Finance*, 29(7), 1751-1767.

Becchetti, L., Ciciretti, R., Dalo, A. & Herzel, S. (2015). Socially Responsible and Conventional Investment Funds: Performance Comparison and the Global Financial Crisis. *Applied Economics*, 47(25), 2541-2562.

Bello, Z.Y. (2005). Socially Responsible Investing and Portfolio Diversification. *Journal of Financial Research,* 28(1), 41–57.

Blake, C., & Morey, M. (2000). Morningstar Ratings and Mutual Fund Performance. *Journal of Financial and Quantitative Analysis*, 35(3), 451-483.

Bogle, J.C. (2010). *The Challenge to Mutual Fund Stewardship* in Haslem, J. A. (ed.) Mutual Funds: Portfolio Structures, Analysis, Management, and Stewardship. John Wiley & Sons, Inc, 283-304.

Bollen, N.P.B. & Busse J.A. (2005). Short-Term Persistence in Mutual Fund Performance. *The Review of Financial Studies*, 18(2), 569–597

Bollen, N.P.B. (2007). Mutual fund attributes and investor behavior. *Journal of Financial and Quantitative Analysis*, 42(3), 683-708.

Boyd, D.P. & Yilmaz, M. (2007). Stewardship As A Factor In The Financial Performance Of Mutual Funds. *Journal of Business & Economic Research*, 5(3), 11-17.

Brenning, C., & Fritzén, V. (2009). The Effect of Morningstar Rating. Working Paper Stockholm School of Economics. Available at: http://arc.hhs.se/download.aspx?mediumid=703.

Benson, K.L., Brailsford, T.J. & Humphrey, J.E. (2006). Do Socially Responsible Fund Managers Really Invest Differently?. *Journal of Business Ethics*, 65(4), 337-357.

Benson, KL. & Humphrey, J. (2008) Socially Responsible Investment funds: Investor reaction to current and past returns. *Journal of Banking and Finance*, 32(9), 1850-1859.

Brown, S., & Goettzmann, W. (1995). Performance Persistence. *The Journal of Finance*, 50(2), 679–698.

Brown, S., Goetzmann, W., Ibbotson, R., & Ross, S. (1992). Survivorship Bias in Performance Studies. *Review of Financial Studies*, 5(4), 553–580.

Cao, X. J., Ghosh, A., Goh, C. Y. J. & Ng, W. S. (2014). Governance Matter: Morningstar Stewardship Grades and Mutual Fund Performance. Research Collection School Of Economics. Available at: http://ink.library.smu.edu.sg/cgi/viewcontent.cgi?article=2718&context=soe_research

Carhart, M. (1997). On Persistence in Mutual Funds Performance. *Journal of Finance*, 52(1), 57-82.

Carhart, M.,Carpenter, J., Lynch, A. & Musto, D. (2002). Mutual Funds Survivorship. *Review of Finance Studies*, 15(5), 1439-1463.

CFA Institute (2015). Environmental, Social, and Governance Issues in Investing: A Guide for Investment Professionals, Available at http://www.cfapubs.org/doi/pdf/10.2469/ccb.v2015.n11.1

Chegut, A., Schenk, H. & Scholtens, B. (2011). Assessing SRI fund performance research: Best practices in empirical analysis. *Sustainable Development*, 19(2), 77–94.

Chen, C. R., & Huang, Y. (2011). Mutual Fund Governance and Performance: A Quantile Regression Analysis of Morningstar's Stewardship Grade. *Corporate Governance: An International Review*, 19(4), 311–333.

Chotivetthamrong, C. (2015). The Performance of Thai Mutual Funds: A 5-Star Morningstar Mutual Fund Rating. *Journal of Economics, Business and Management*, 3(1),104-110.

Chou,W.-H., Ng, L. & Wang, Q. (2011). Are better governed funds better monitors? funds?, *Journal of Corporate Finance*, 17(5), 1254–1271.

Cortez, M.C., Silva, F. & Areal, N. (2011). The performance of european socially responsible funds. *Journal of Business Ethics*, 87(4), 573–588.

Cortez, M.C., Silva, F. & Areal, N. (2012). Socially responsibe investing in the global market: The performance of US and European funds. International. *Journal of Finance and Economics*, 17(3), 254–271.

Cremers, M., Driessen, J., Maenhout, P. & Weinbaum, D. (2009). Does skin in the game matter? Director incentives and governance in the mutual fund industry. *Journal of Financial and Quantitative Analysis*, 44(6), 1345-1373.

Das, P. K. & Uma Rao, S.P. (2014). Performance persistence in socially responsible mutual funds. *International Journal of Economics and Business Research*, 8(4), 490–503.

Del Guercio, D., Dann, L. Y. & Partch, M. M. (2003). Governance and boards of directors in closed-end investment companies. *Journal of Financial Economics*, 69(1), 111-152.

Del Guercio, D., & Tack, P. A. (2008). Star Power: The Effect of Morningstar Ratings on Mutual Fund Flow. *Journal of Financial and Quantitative Analysis*, 43(4), 907-936.

DiBartolomeo, D. & L. Kurtz, 1999. Managing Risk Exposures of Socially Screened Portfolios, Northfield Information Services, Boston, Available at www.northinfo.com.

Ding, B. & Wermers, R. (2012). Mutual Fund Performance and Governance Structure: The Role of Portfolio Managers and Boards of Directors. Available at SSRN: https://ssrn.com/abstract=2207229

Duret, A.-S., Hereil, P., Mitaine, P., Moussavi, N., & Roncalli, T. (2008). Fund rating systems and performance predictability. Working Paper. University of Evry. Available at SSRN: http://ssrn.com/abstract=1121383.

El Ghoul, S. & A. Karoui (2017). Does corporate social responsibility affect mutual fund performance and flows?. *Journal of Banking and Finance*, 77, 53-63.

Efron, B. (1979). Bootstrap methods: another look at the jackknife. *Annals of Statistics*, 7(1), 1-26.

Evans, A. L. (2008). Portfolio manager ownership and mutual fund performance. Financial Management, 37(3), 513–534.

Faff, R., Parwada, J., & Poh, H. (2007). The Information Content of Australian Managed Fund Ratings. *Journal of Business Finance & Accounting*, 34 (9-10), 1528-1547.

Ferson, W., & Schadt, R.W. (1996). Measuring Fund Strategy and Performance in Changing Economic Conditions. *The Journal of Finance*, 51(2), 425-461.

Flammer, C. (2015). Does Corporate Social Responsibility Lead to Superior Financial Performance? A Regression Discontinuity Approach. *Management Science*, 61(11), 2549-2568.

Friede, G, Busch, T. & Bassen, A. (2015). ESG and financial performance: aggregated evidence from more than 2000 empirical studies, *Journal of Sustainable Finance & Investment*, 5(4), 210–233.

Ferris, S. P. & Xuemin, Y.S. (2007). Do independent directors and chairmen really matter? The role of boards of directors in mutual fund governance. *Journal of Corporate Finance*, 13(2-3), 392-420.

Füss, R., Hille, J., Rindler, P., Schmidt, J. & Schmidt, M. (2010). From Rising stars and falling angels: on the relationship between the performance and ratings of German mutual funds. *Journal of Wealth Management*, 13(1), 75-90.

Garnier, O., & Pujol, T. (2007). Can today's stars be used to read the stars of the future? Case study of equity funds distributed in France between 2000 and 2005, Working Papers Autorité des marchés financiers. Available at http://www.amf-france.org/technique/multimedia?docId=workspace://SpacesStore/b1fae1f7-a61f-4bb1-b7ee-5f920d2a99ca_en_1.0_rendition.

Gerrans, P. (2006). Morningstar ratings and future performance. *Accounting & Finance*, 46(4), 605-628.

Goh, J., Ghosh, A. & Seng, W. (2012). Grades Matter in Performance: Morningstar Stewardship Grades and Mutual Fund Performance. Finance and Financial Manegement Commons. Available at: http://ink.library.smu.edu.sg/lkcsb_research/3324

Goldreyer, E. & Diltz, D. (1999). The Performance of Socially Responsible Mutual Funds: Incorporating Sociopolitical Information in Portfolio Selection. *Managerial Finance*, 25(1), 23-36.

Gottesmam, A. & Morey, M. (2012). Mutual Fund corporate culture and performance. *Review of Financial Economics*, 21(2), 69-81.

Gregory, A., Matatko, J., & Luther, R. (1997). Ethical until trust financial performance: Small company effects and fund size effects. *Journal of Business Finance and Accounting*, 24(5), 705–725.

Gregory A & Whittaker, J.M. (2007). Performance and performance persistence of ethical unit trusts in the UK. *Journal of Business Finance and Accounting*, 34(7-8), 1327–1344.

Grinblatt, M., & Sheridan, T. (1992). The Persistence of Mutual Fund Performance. *Journal of Finance*, 47(5),1977–1984.

Grinblatt, M., Titman, S. & Wermers, R. (1995). Momentum investment strategies, portfolio performance, and herding: A study of mutual fund behaviour. *American Economic Review*, 85(5), 1088–1105.

Grinblatt, M. & Keloharju, M. (2000). The investment behavior and performance of various investor types: a study of Finland's unique data set. *Journal of Financial Economics*, 55(1), 43-67.

Gruber, M. (1996). The growth in actively managed mutual funds. *The Journal of Finance*, 51 (3), 783–810.

GSIA (2017): 2016 Global sustainable Investment Review, Available at https://www.msci.com/esg-fund-metrics

Hamilton, S., Jo, H. & Statman, M. (1993). Doing Well While Doing Good? The Investment Performance of Socially Responsible Mutual Funds. *Financial Analysts Journal*, 49(6), 62-66.

Haslem, J. A. (2014). Morningstar Mutual Fund Measures and Selection Model. *Journal of Wealth Management*, 17(2), 19-30.

Hendricks, D., Patel, J. & Zeckhauser, R. (1993). Hot hands in Mutual Funds: Short-Run Persistence of Relative Performance. *Journal of Finance*, 48(1), 93-130.

Hazenberg, Z.Z. (2012). Investment Fund Governance: An empirical investigation of Luxembourg UCITS. Thesis at University of Luxembourg. Available at http://orbilu.uni.lu/handle/10993/15562

Hill, R.P., Ainscough, T., Shank, T. & D. Manullang (2007). Corporate social responsibility and social responsible investing: A global perspective. *Journal of Business Ethics*, 70(2), 165–174.

Howe, T., & Pope, R. (1996). Equity Mutual Fund Historical Performance Ratings. *Journal of Financial And Strategic Decisions*, 9(1), 33-43.

Humphrey, J.E., Warren, G.J. & Boon, J. (2016). What is Different about Socially Responsible Funds? A Holdings-Based Analysis. *Journal of Business Ethics*, 138(2), 263-277.

Junkus, J. & Berry, T. D. (2015). Socially responsible investing: a review of the critical issues. *Managerial Finance*, 41(11), 1176-1201.

Jun, X., Li, M., Yan, W., & Zhang, R. (2014). Flow-Performance Relationship and Star Effect: New Evidence from Chinese Mutual Funds. *Emerging Markets Finance and Trade*, 50(3), 81-101.

Kamal, R. (2013). Can Morningstar Analyst Ratings Predict Fund Performance?. *The Journal of Applied Business Research*, 29(6), 1665-1672.

Kelly, B. & Jiang, H. (2014). Tail Risk and Asset Prices. *The Review of Financial Studies*, 27(10), 2841-2871.

Khorana, A., Servaes, H. & Wedge, L. (2007). Portfolio manager ownership and fund performance. *Journal of Financial Economics*, 85(1), 179–204.

Khorana, A., Tufano P. & Wedge, L. (2007). Board structure, merger and shareholder wealth.: a study of mutual fund industry. *Journal of Financial Economics*, 85(2), 571-598.

Koenker, R., & Basset, G. (1978). Regression Quantiles. *Econometrica*, 46(1), 33-50.

Kong, S. X. & Tang, D. Y. (2008). Unitary boards and mutual fund governance. *Journal of Financial Research*, 31(3), 193-224.

Kreander, N., Gray, R.H., Power, D.M. & Sinclair, C.D. (2002). The Financial Performance of European Ethical Funds 1996-1998, *Journal of Accounting and Finance*, 1, 3-22.

Kreander, N., Gray, R.H., Power, D.M. & Sinclair, C.D. (2005). Evaluating the Performance of Ethical and Non-ethical Funds: A Matched Pair Analysis. *Journal of Business Finance and Accounting*, 32(7-8), 1465-1493.

Kryzanowski, L. & Mohebshahedin, M. (2016). Board governance, monetary interest, and closed – end fund performance. Journal of Corporate Finance, 38, 196-217.

Kurtz, L. (1997). The Impact of Social Screening on Growth-Oriented Investment Strategies. The Journal of Performance Measurement, Spring, 65-71.

Lai, S., Tiwari A. & Zhang, Z. (2010). Mutual Fund Flows, Performance Persistence, and Board Quality. Available at SSRN: https://ssrn.com/abstract=1570525

Lean, H.H., Ang, W.R. & Smyth, R. (2014). *Performance persistence of socially responsible investment funds in the Asia Pacific region* in G. Gregoriou and L. David (Eds.), Handbook of Asian finance REITs, trading, and fund performance, Academic Press Inc, Oxford, 377–392.

Lean, H.H., Ang, W.R. & Smyth, R. (2015). Performance and performance persistence of socially responsible investment funds in Europe and North America. *The North American Journal of Economics and Finance*, 34, 254–266.

Leite, P. & Cortez, M.C. (2014). Style and performance of international socially responsible funds in Europe. *Research in International Business and Finance*, 30(1), 248–267.

Leite, P. & Cortez, M.C. (2015). Performance of European Socially Responsible Funds during Market Crises: Evidence from France. *International Review of Financial Analysis*, 40, 132-141

Lewis, A., & Mackenzie, C. (2000). Morals, Motives and Money: The Case of U.K. Ethical Investing. *Human Relations*, 53(2): 179–191.

Luther, R.G., Matatko, J. & Corner, D.C. (1992). The Investment Performance of UK 'Ethical' Unit Trusts'. *Accounting, Auditing and Accountability Journal*, 5(4), 57-70.

Mallin, C.A., Saadouni, B. & Briston, R.J. (1995). The financial performance of ethical investment funds. *Journal of Business Finance and Accounting*, 22(4), 483-496.

Mamatzakis, E. & Xu, B. (2017). Does corporate governance matter in fund management company?: the case of China, Available at https://mpra.ub.uni-muenchen.de/76138/1/MPRA_paper_76138.pdf

Margolis, J.D., Elfenbein, H.A., & Walsh, J.P. (2009). Does it pay to be good? A meta-analysis and redirection of research on the relationship between corporate social and financial performance. Available at SSRN: https://ssrn.com/abstract=1866371.

Meinhardt, C. (2014). Ratings and Performance of German Mutual Funds – A Comparison of Feri Trust, Finanztest and FondsNote. Available at SSRN: http://ssrn.com/abstract=1943577.

Meschke, F. (2007). An empirical examination of mutual fund boards. Available at SSRN: http://ssrn.com/abstract=676901

Morey, M. (2002). Mutual Fund Age and Morningstar Ratings. *Financial Analysts Journal*, 58 (2), 56-63.

Morey, M. (2005). The Kiss Of Death: A 5-Star Morningstar Mutual Fund Rating?. *Journal of Investment Management*, 3(2), 41-52.

Morey, M., & Gottesman, A. (2006). Morningstar Mutual Fund Ratings Redux. *Journal of Investment Consulting*, 8(1), 25–37.

Morningstar (2010). The Morningstar Stewardship Grade for Funds. Available at http://corporate.morningstar.com/us/documents/stewardshipgradefunds/StewGradeMethodology_06-18-07.pdf

Morningstar (2011). Morningstar Analyst Rating for Funds Methodology Document. Available at https://corporate.morningstar.com/us/documents/MethodologyDocuments/AnalystRatingforFundsMethodology.pdf

Morningstar (2016a). The Morningstar Sustainable Investing Handbook. Available at http://corporate1.morningstar.com/Morningstar-Sustainable-Investing-Handbook.pdf

Morningstar (2016b). The Morningstar Sustainability Rating. Available at http://corporate1.morningstar.com/Morningstar-Sustainable-Rating-Methodology.pdf

Müller, S., & Weber, M. (2014). Evaluation the Rating of Stiftung Warentest: How good are the Ratings and can they be improved?. *European Financial Management*, 20(2), 207-235.

Muñoz, F., Vargas, M. & Marco, I. (2014). Environmental Mutual Funds: Financial Performance and Managerial Abilities. *Journal of Business Ethics*, 124(4), 551–569.

Nanigian, D. (2012). Why do mutual funds expenses matter?. *Financial Services Review*, 21(3), 239 –257.

Ng, W.S. (2009). Does Morningstar Shine in the Universe of Mutual Funds? A Study on Morningstar Mutual Fund Ratings. Singapore Management University Master Thesis, Available at: http://ink.library.smu.edu.sg/etd_coll/10

Nofsinger, J. & Sias, R. (1999). Herding and Feedback Trading by Institutional and Individual Investors. *Journal of Finance*, 54(6), 2263-2295.

Nofsinger, J. & Varma, A. (2014). Socially responsible funds and market crises. Journal of Banking and Finance, 48, 180–193.

Philips, C., & Kinniry, F. (2010). Mutual fund ratings and future performance. Vanguard, 1-12, Available at http://www.vanguard.com/pdf/icrwmf.pdf.

Qian, M. (2006). Whom Can You Trust? A Study on Mutual Fund Governance. Available at SSRN: https://ssrn.com/abstract=685543

Renneboog, L., Ter Horst, J. & C. Zhang. (2008a). The price of ethics and stakeholder governance: The performance of socially responsible mutual funds. *Journal of Corporate Finance*, 14(3), 302–322.

Renneboog, L., Ter Horst, J & C. Zhang (2008b). Socially responsible investments: Institutional aspects, performance, and investor behavior. *Journal of Banking and Finance*, 32(9), 1723–1742.

Renneboog, L, Horst, J & Zhang, C. (2011). Is Ethical Money Financially Smart? Nonfinancial Attributes and Money Flows of Socially Responsible Investment Funds. *Journal of Financial Intermediation*, 20(4), 562-588.

Rudd, A. (1981). Social Responsibility and Portfolio Performance. *California Management Review*, 23(4), 55–61.

Sauer, D. (1997). The Impact of Social-Responsibility Screens on Investment Performance: Evidence from the Domini 400 Social Index and Domini Equity Fund, Review of Financial Economics, 6(2), 137-149.

Scholtens, B. (2005). Style and performance of Dutch social responsible investment funds. *Journal of Investing*, 14(1), 63-72.

Statman, M. & Glushkov, D. (2016). Classifying and Measuring the Performance of Socially Responsible Mutual Funds. *Journal of Portfolio Management*, 42(2), 140-151

Statman, M. (2000). Socially Responsible Mutual Funds. *Financial Analysts Journal*, 56(3), 30-39.

Sustainalytics (2016). Sustainalytics' Research Methodology. Company ESG Research.

Trahan, E.A. (2008). Mutual fund governance and fund performance. *Corporate Ownership & Control*, 5(4),384-392.

Tufano, P. & Sevick, M. (1997). Board structure and fee-setting in the US mutual fund industry. *Journal of Financial Economics*, 46(3), 321–355.

Vidal-García, J. (2013). The persistence Of European Mutual Fund Performance. *Research in International Business and Finance*, 28, 45-67.

Webley, P., Lewis, A. & Mackenzie, C. (2001). Commitment Among Ethical Investors: An Experimental Approach. *Journal of Economic Psychology*, 22(1), 27-42.

Wellman, J. W. & Zhou, J. (2008). Corporate Governance and Mutual Fund Performance: A First Look at the Morningstar Stewardship Grades. Available at SSRN: https://ssrn.com/abstract=1107841

Wermers, R. (1999). Mutual Fund Herding and the Impact on Stock Prices, Journal of Finance, 54(2), 581-622.

Wimmer, M. (2012). ESG-Persistence in Socially Responsible Mutual Funds. Available at SSRN: https://ssrn.com/abstract=2138882

White, M. (1995). The Performance of Environmental Mutual Funds in the United States and Germany: Is there Economic Hope for "Green" Investors?. Research in Corporate Social Performance and Policy, Supplement 1, 323- 344.

APENDDIX

Table 37- Cross sectional Yearly Alpha Ratio

Variable	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014
l1.stars4	0.3465	10.546	-2.6092**	0.7886	-0.1326	-0.0639	-1.3805*	-1.3743**	-1.7699***	-1.6551***	-0.9953***
l1.stars3	-2.0737*	0.0908	-12.671	1.3468*	-14.491	-0.6517	-1.7511**	-1.7412***	-3.1438***	-3.1120***	-1.8293***
l1.stars2	-2.4073*	0.5101	-0.6558	1.8568**	-1.9612*	-0.3558	-1.7153**	-2.3797***	-3.5973***	-4.4160***	-2.7856***
l1.stars1	-4.1337***	-0.1187	-3.4873***	-0.4293	-17.086	0.0717	-2.5637**	-4.1585***	-4.9201***	-6.1315***	-4.3172***
Largecapblend	0.9609	1.4079*	-15.676	-2.0834**	13.347	3.7379***	2.8036***	-0.9174	0.4702	-0.4582	0.1917
Largecapgrwth	23.104	0.2374	18.409	-17.055	4.4238**	3.4828*	3.2692*	-13.669	-0.2775	-0.605	-0.4909
Largecapvalue	-0.8598	0.0889	-2.2598*	0.7169	4.2876***	2.7899**	2.4619**	-1.7742**	0.6739	-1.3897*	0.4641
Midcap	0.7193	0.3748	-2.8022*	-5.5996***	23.929	4.5212***	3.0099*	-16.036	15.813	0.0789	0.1248
Smallcap	15.652	2.6694**	-0.6278	1.13	5.6740***	3.7871**	4.1412***	-2.5002***	1.5769*	0.2854	0.2347
LargecapexUK	10.493	1.7596*	-0.5128	-0.3624	2.7413*	2.8660**	2.9389***	-1.5364**	0.461	-12.437	0.6278
SmallmidexUK	-93.953	-1.356	1.603	-2.6657***	8.8624*	0.998	29.355	-3.8539***	0.6825	-21.461	0.486
Eurozoneflexcap	-82.692	11.421	-49.953	-0.5315	0.7524	4.3149**	2.178	0.8248	0.3369	0.3386	1.2384*
Eurozonelargecap	0.7913	0.8467	-13.781	-1.5495**	2.5217*	3.9301***	3.1893***	-1.0458*	0.9951	-0.0717	0.4496
Eurozonemidcap	1.427	0.6855	-28.821	-11.041	22.992	3.202	20.259	0.4654	-13.045	0.3059	2.8614***
Eurozonesmallcap	-14.2839**	7.1779**	-10.1311**	-43.192	-12.504	20.901	42.101	-0.8958	0.8402	11.521	-0.6671
_cons	-0.5581	-1.6639*	24.779	0.9535	-0.9151	-3.2951**	-16.555	3.0024***	1.7408**	2.7275***	1.0554**
N	352	405	473	514	667	778	911	1029	1145	1241	1345
r2	0.1988	0.0672	0.105	0.0675	0.0635	0.0407	0.0324	0.0624	0.0748	0.1242	0.0956
F	36.165	19.334	35.968	32.748	24.655	14.336	1.364	35.124	46.368	97.084	75.243

This table reports the coefficients for Cross sectional Yearly models for Alpha L1.Star is the one year lagged variable representing the rating of the mutual fund and yr* are the time dummies variables and, finally, Largecapblend, Largecapgrwth, Largecapvalue, Midcap, Smallcap, LargecapexUK, SmallmidexUK, Eurozoneflexcap, Eurozonelargecap, Eurozonemidcap and Eurozonesmallcap are dummies to control for categories. N is the number of observations, r2 the pseudo-squared fit measure, Rho is the fraction of variance due to individual effects and Chi-square (p) is the p-value associated to the Chi-square significance test. *Significant at 10%; ** significant at 5% and * significant at 1%.**

Table 38- Cross sectional Yearly Sharpe Ratio

Variable	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014
stars4	-0.1610*	-0.2085***	-0.3038***	-0.0531	-0.0461**	-0.0829***	-0.1108***	-0.1725***	-0.2141***	-0.1420***	-0.2309***
stars3	-0.3343***	-0.4242***	-0.3818***	-0.1388***	-0.1184***	-0.1323***	-0.1784***	-0.2538***	-0.3597***	-0.2875***	-0.4019***
stars2	-0.4028***	-0.5052***	-0.5532***	-0.1619***	-0.1843***	-0.1854***	-0.2383***	-0.3071***	-0.4625***	-0.4254***	-0.6129***
stars1	-0.4361***	-0.5878***	-0.9026***	-0.4209***	-0.2862***	-0.2644***	-0.3137***	-0.4251***	-0.6081***	-0.5774***	-0.8425***
Largecapblend	-0.2881**	-0.2376**	-0.0718	0.0599	-0.0777***	-0.0431	-0.1270***	-0.0614*	-0.0473	0.0055	-0.0639
Largecapgrwth	-0.4270***	-0.2553**	0.0111	0.2502***	0.0418	0.0159	0.0023	0.1883***	0.2163***	0.1246**	0.0323
Largecapvalue	-0.3281**	-0.2873**	-0.0895	0.0167	-0.0415	-0.0265	-0.1505***	-0.2046***	-0.1603***	-0.0866***	-0.2378***
Midcap	0.0301	0.4201***	0.2372**	-0.0014	0.0939***	0.0145	0.1308***	0.1991***	0.1992***	0.0596	0.1965***
Smallcap	0.1428	0.4887***	0.2778**	0.0594	0.0494	-0.0053	0.1162***	0.2886***	0.2101***	0.1208***	0.2392***
LargecapexUK	-0.21	-0.0869	0.0975	0.3015***	0.1076***	0.0423	-0.0881***	-0.1774***	-0.0522	0.0550*	0.1526***
SmallmidexUK	0.0064	-0.0036	0.241	0.1172	0.2622***	0.1699***	0.1955***	0.0608	0.1791**	-0.0399	0.2676***
Eurozoneflexcap	-0.0729	-0.1664	-0.1458	-0.1561	0.024	0.0104	-0.1093**	-0.3012***	-0.2605***	-0.1839***	-0.2345***
Eurozonelargecap	-0.2091	-0.2826**	-0.1794**	0.1366***	0.0585**	0.0423	-0.2077***	-0.4000***	-0.3208***	-0.1399***	-0.2861***
Eurozonemidcap	0.1976	0.1366	-0.1167	0.0261	0.1661***	0.0713	0.0212	-0.0573	-0.0424	-0.0636*	0.0096
Eurozonesmallcap	-0.2733	-0.219	-0.4042***	-0.2577*	0.0166	0.0417	0.0582	0.1284*	-0.0098	-0.0338	-0.089
_cons	0.1402	1.7416***	2.1054***	1.1975***	-0.7648***	-0.4515***	-0.1105***	0.7418***	0.6914***	0.8164***	1.8866***
N	347	399	466	507	659	769	902	1020	1137	1233	1337
r2	0.3619	0.5743	0.3939	0.1873	0.362	0.2712	0.4698	0.536	0.5668	0.4725	0.5191
F	17.7361	32.383	26.9785	8.4156	26.1943	18.1597	48.0609	76.9651	112.4071	74.7891	93.9336

This table reports the coefficients for Cross sectional Yearly models for Sharpe Ratio. L1.Star is the one year lagged variable representing the rating of the mutual fund and yr* are the time dummies variables and, finally, Largecapblend, Largecapgrwth, Largecapvalue, Midcap, Smallcap, LargecapexUK, SmallmidexUK, Eurozoneflexcap, Eurozonelargecap, Eurozonemidcap and Eurozonesmallcap are dummies to control for categories. N is the number of observations, r2 the pseudo-squared fit measure, Rho is the fraction of variance due to individual effects and Chi-square (p) is the p-value associated to the Chi-square significance test. *Significant at 10%; ** significant at 5% and * significant at 1%.**

Table 39- Cross sectional Yearly Return models

Variable	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014
l1.stars4	0.7335	1.6293	-2.4431**	0.8042	0.7319	2.2895**	-1.3259	-2.5281***	-0.7471	0.388	-1.0090***
l1.stars3	-1.4163	0.4328	-3.3198***	1.3370*	0.1859	2.8690***	-1.8637**	-3.2992***	-1.8471***	0.4583	-1.9942***
l1.stars2	-1.3114	1.1088	-4.8222***	1.7364**	1.1811	3.6653***	-1.8536**	-3.8339***	-2.3018***	-0.338	-2.9877***
l1.stars1	-2.3680*	0.4264	-7.1562***	-0.3202	2.7388**	4.5570***	-2.7818**	-6.5673***	-3.4163***	-1.6670**	-4.2693***
Largecapblend	-4.7312***	-2.2261*	-2.6686***	1.0549	0.6192	-3.8326**	-6.2773***	3.5520***	-1.2996*	-3.0145***	0.3428
Largecapgrwth	-5.5211***	-2.2545	-2.2082**	3.0533*	3.5769**	-2.5892	1.4701	5.3799***	2.1375*	-5.0105***	0.921
Largecapvalue	-4.5560**	-4.1675***	-2.2491*	1.0513	2.2630*	-5.7626***	-9.8041***	1.6714**	-1.9416**	-1.1895	-1.1311*
Midcap	6.4307***	4.0762**	7.9774***	-3.3401**	-2.0352	7.8834***	8.5784***	-1.9007*	6.0698***	2.1037*	1.0117
Smallcap	6.8630***	8.7397***	9.5096***	-2.0752	-5.1122***	11.6623***	13.1273***	-2.8697***	5.9355***	8.0255***	0.2104
LargecapexUK	-1.8079	2.2764*	0.9584	3.9411***	3.3522***	-4.5163**	-6.0746***	1.1895	4.1295***	0.6479	1.8342***
SmallmidexUK	-2.1378	3.9267	8.0899*	-0.7853	2.2172	10.7761	11.8432***	-4.6959***	4.0167*	4.0394	1.0661
Eurozoneflexcap	-6.3226	-0.7626	-4.7742	0.1973	0.1172	-3.6653	-7.6905***	-4.9109***	-0.1656	0.3059	-1.8129**
Eurozonelargecap	-4.4046**	-4.0633***	-2.9421***	5.1678***	2.7376***	-7.4670***	-14.6624***	-0.8968	-0.0237	-0.6841	-1.8558***
Eurozonemidcap	5.3695***	2.6055	7.7454***	-0.4488	-0.2028	6.2361*	3.3313	-4.0323***	2.5656	3.2664**	-0.172
Eurozonesmallcap	-4.4757	12.7615***	-4.6467**	-3.5651	-3.1534	6.9809	4.7086	-3.4885**	0.165	4.3030***	-1.4365
_cons	14.2087***	26.2073***	23.6132***	-0.8332	-45.2903***	31.2747***	18.9595***	-11.6638***	20.2903***	22.3123***	6.3030***
N	352	405	473	514	667	778	911	1029	1145	1241	1345
r2	0.421	0.3409	0.4123	0.1921	0.109	0.2695	0.5515	0.2883	0.1874	0.1969	0.1731
F	23.0922	13.6175	19.7274	12.6096	6.0314	18.3487	60.7363	24.1035	18.343	16.9468	21.2588

This table reports the coefficients for Cross sectional Yearly models for Yerly return. L1.Star is the one year lagged variable representing the rating of the mutual fund and yr* are the time dummies variables and, finally, Largecapblend, Largecapgrwth, Largecapvalue, Midcap, Smallcap, LargecapexUK, SmallmidexUK, Eurozoneflexcap, Eurozonelargecap, Eurozonemidcap and Eurozonesmallcap are dummies to control for categories. N is the number of observations, r2 the pseudo-squared fit measure, Rho is the fraction of variance due to individual effects and Chi-square (p) is the p-value associated to the Chi-square significance test. *Significant at 10%; ** significant at 5% and * significant at 1%.**

.

Table 40- Estimates using three years rating instead overall.

Variable	Alpha	Sharpe	Return
l1.3ystars4	-0.7798***	-0.1571***	-0.4791**
L1.3ystars3	-1.1767***	-0.2583***	-0.9517***
L1.3ystars3	-1.3146***	-0.3556***	-1.1545***
L1.3ystars1	-1.5152***	-0.4781***	-1.7681***
yr2004c	-1.1605***	-1.7457***	6.1924***
yr2005c	0.7420*	-0.1822***	21.5183***
yr2006c	-1.2666***	0.2434***	14.8062***
yr2007c	1.5963***	-0.2397***	-1.8390***
yr2008c	5.2727***	-2.2723***	-47.4104***
yr2009c	3.2191***	-1.9663***	26.5923***
yr2010c	3.7277***	-1.7765***	6.7931***
yr2011c	-0.2034	-1.0179***	-17.8200***
yr2012c	3.9849***	-1.1403***	14.8913***
yr2013c	3.8815***	-0.8915***	17.4675***
Largecapblend	-1.1804***	-0.0629***	-1.4227***
Largecapgrwth	0.727	0.0686***	0.1201
Largecapvalue	-1.5970***	-0.1515***	-2.1838***
Midcap	-0.0223	0.1447***	3.1576***
Smallcap	0.3243	0.1807***	4.5674***
LargecapexUK	0.0065	0.0173	0.5398
SmallmidexUK	1.8610*	0.1518***	3.2813***
Eurozoneflexcap	-2.5954***	-0.1777***	-2.2866***
Eurozonelargecap	-1.0720***	-0.2004***	-2.7231***
Eurozonemidcap	2.1813*	0.0103	1.7368**
Eurozonesmallcap	1.8061*	-0.0351	0.8201
_cons	-0.1827	1.7229***	5.9386***
N	4435	8776	8860
r2_w	0.1752	0.9353	0.9185
rho	0.0699	0.2067	0.0000

***Significant at 10%; ** significant at 5% and *** significant at 1%.**